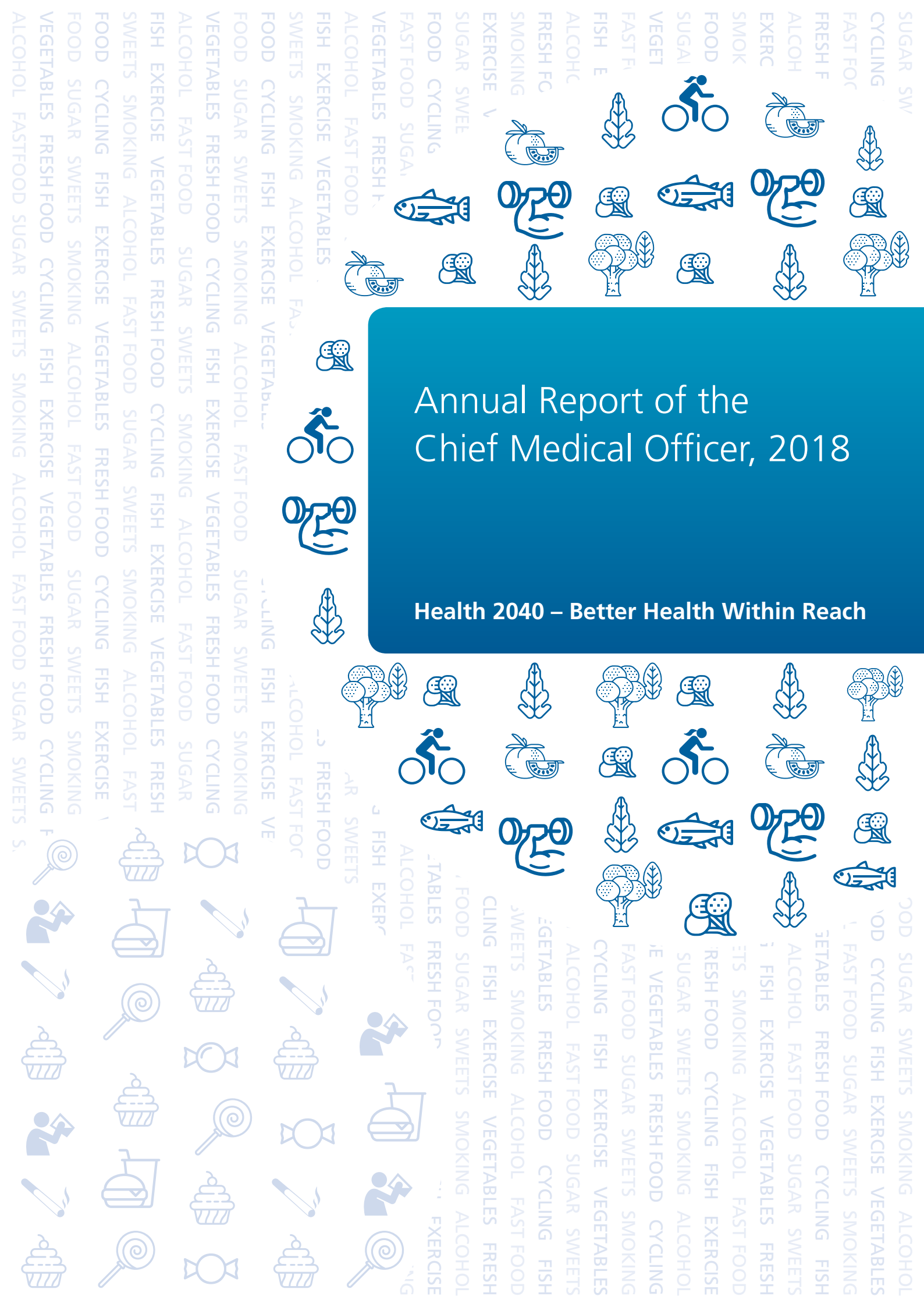
Annual Report of the Chief Medical Officer, 2018

Health 2040 – Better Health Within Reach



Discussion of health often focuses on the financial cost of healthcare. Far from a cost, our health is our primary asset, as individuals, communities and as a nation. Maintaining 'good health' and preventing 'ill-health' is an investment for the future. This is so important that we need to measure and track health in our society. After all, as Peter Drucker said, "What gets measured gets managed."

I look at the future health of England's population in this, my tenth, annual report. As the NHS has been developing its own 10-year plan, I look further ahead. I wanted this report to take an aspirational view of what health could and should look like in 2040 if we commit to it being our nation's primary asset.

Every part of the health system has a role to play in creating a healthier and fairer future. The fortunate truth is that we already know how to make fantastic improvements and prepare for better health that is 'within our reach'. The green shoots of a brighter future are already visible in some parts of our health system. Now we need to develop, plan and scale, harnessing technology (including wearables and AI) to support this.

We need to develop our environment to make the healthy choice the easy choice, thus promoting our health, our happiness and our economy whilst preventing disease.

I hope this report inspires all readers to understand that we can achieve better health in England in 2040 – this can be our shared vision, with each of us delivering our part, in our different ways.

Prof Dame Sally C Davies

Editors and authors

Editors and authors

This report could not have been produced without the generous input of the following people.

Editor-in-Chief

Jonathan Pearson-Stuttard, Imperial College London

Project Manager and Editor

Orla Murphy, Department of Health and Social Care

Chapter Authors

Chapter 1 Chief Medical Officer's summary

Chapter lead and author
Sally C Davies, Chief Medical Officer

Chapter 2 Health and economic outcomes

Chapter authors
Paul Johnson, Institute for Fiscal Studies
George Stoye, Institute for Fiscal Studies
David Sturrock, Institute for Fiscal Studies

Chapter 3 The local health environment

Chapter lead
Tim Elwell-Sutton, The Health Foundation

Chapter authors
Tim Elwell-Sutton, The Health Foundation
Louise Marshall, The Health Foundation
David Finch, The Health Foundation
Jo Bibby, The Health Foundation

Chapter 4 Social health

Chapter lead and author
Tom Kibasi, Institute for Public Policy Research

Chapter 5 How will health be experienced in 2040?

Chapter lead
Martin Stewart-Weeks, Public Purpose Pty Ltd

Chapter authors
Martin Stewart-Weeks, Public Purpose Pty Ltd
Junaid Bajwa, MSD *and* Imperial College London
Mobasher Butt, Sloane Medical Practice *and* GP at Hand (NHS) *and* Babylon Healthcare Ltd

Chapter 6 Demography

Chapter lead
Lucinda Hiam, London School of Hygiene and Tropical Medicine

Chapter authors
Lucinda Hiam, London School of Hygiene and Tropical Medicine
Martin McKee, London School of Hygiene and Tropical Medicine
Danny Dorling, University of Oxford

Chapter 7 Multimorbidity

Chapter lead
Chris Whitty, Department of Health and Social Care *and* London School of Hygiene and Tropical Medicine

Chapter authors
Chris Whitty, Department of Health and Social Care *and* London School of Hygiene and Tropical Medicine
Alexandra Lee, Department of Health and Social Care

Special section – Mental health disorder

Section Author
Anto Ingrassia, Children and Young People's Health Partnership, Guy's And St Thomas' NHS Foundation Trust

Chapter 8 Changing behaviour for a healthier population

Chapter lead
Theresa M Marteau, Behaviour and Health Research Unit, University of Cambridge

Chapter authors
James G McGowan, THIS Institute (The Healthcare Improvement Studies Institute), University of Cambridge
Mark Petticrew, London School of Hygiene and Tropical Medicine
Harry Rutter, Department of Social and Policy Sciences, University of Bath
Jonathan Pearson-Stuttard, Department of Health and Social Care *and* Imperial College London
Martin White, University of Cambridge and Public Health Research Programme, NIHR *and* UK Society for Behavioural Medicine
Theresa M Marteau, Behaviour and Health Research Unit, University of Cambridge

Chapter 9 Health inequalities – a challenge to current health policy

Chapter lead
Majid Ezzati, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London *and* MRC-PHE Centre for Environment and Health, Imperial College London *and* WHO Collaborating Centre on NCD Surveillance and Epidemiology, Imperial College London

Chapter authors
James Bennett, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London *and* MRC-PHE Centre for Environment and Health, Imperial College London

Jonathan Pearson-Stuttard, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London *and* MRC-PHE Centre for Environment and Health, Imperial College London *and* Department of Public Health and Policy, University of Liverpool, Liverpool, UK

Vasilis Kontis, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London *and* MRC-PHE Centre for Environment and Health, Imperial College London, London, UK

Simon Capewell, Department of Public Health and Policy, University of Liverpool

Ingrid Wolfe, Department of Primary Care and Public Health Sciences, Kings College London *and* Evelina London Child Health Partnership, Evelina London Children's Healthcare, Guy's and St Thomas' NHS Trust

Richard Blundell, Department of Economics, University College London *and* ESRC Centre for the Microeconomic Analysis of Public Policy, Institute for Fiscal Studies

Majid Ezzati, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London *and* MRC-PHE Centre for Environment and Health, Imperial College London *and* WHO Collaborating Centre on NCD Surveillance and Epidemiology, Imperial College London

Chapter 10 Machine learning for individualised medicine

Chapter lead
Mihaela van der Schaar, University of Cambridge *and* The Alan Turing Institute

Chapter authors
Mihaela van der Schaar, University of Cambridge *and* The Alan Turing Institute

William Zame, Departments of Economics and Mathematics, University of California, Los Angeles

Chapter 11 Emerging technologies – population scale impacts

Chapter lead and author
Maurizio Vecchione, Global Good and Research, Intellectual Ventures

Chapter 12 Emerging technologies in healthcare

Chapter lead
Dominic King, DeepMind *and* Imperial College London

Chapter authors
Dominic King, DeepMind *and* Imperial College London

Alan Karthikesalingam, DeepMind *and* Imperial College London

Geraint Rees, DeepMind

Chapter 13 Data, technology, trust and fairness

Chapter authors
Matt Fenech, Future Advocacy

Olly Buston, Future Advocacy

Mike Parker, Wellcome Centre for Ethics and Humanities, University of Oxford

Chapter 14 Embracing uncertainty: futures thinking in action

Chapter lead
Jonathan Grant, The Policy Institute, King's College London

Chapter authors
Harriet Boulding, The Policy Institute, King's College London

Hugo Harper, The Behavioural Insights Team

Ross Pow, Power of Numbers

David Halpern, The Behavioural Insights Team

Jonathan Grant, The Policy Institute, King's College London

Chapter 15 Forecasts for health in 2040

Chapter authors
Stein Emil Vollset, Institute for Health Metrics and Evaluation

Christopher J L Murray, Institute for Health Metrics and Evaluation

Chapter 10

Machine learning for individualised medicine

Chapter lead
Mihaela van der Schaar[1,2]

Chapter authors
Mihaela van der Schaar[1,2], William Zame[3]

1 University of Cambridge
2 The Alan Turing Institute
3 University of California Los Angeles

01 Machine learning in healthcare: challenges and opportunities

In 2015, Eric Topol[1] presented persuasive evidence and arguments that current advances in medical technology – and further advances that are likely to come in the very near future (including wearable devices, faster and cheaper means of genomic sequencing) – will enormously increase the amount of data that is available for individual patients. But as Topol points out, if we wish to use this data to predict, prevent and treat illness – especially chronic illness – data will not be enough: "we want answers, not just data." In particular, we need predictive power. And as Topol also points out, "such predictive power must rely on machine learning."

By machine learning (ML) we mean the process by which computer systems can learn directly from data, examples and experience, rather than being taught on the basis of pre-determined rules.* The purpose of this chapter is to illustrate some of the progress ML has already made in healthcare and to suggest some of the progress it might make – and ought to make – in the near future. The view presented here is deliberately optimistic; for ML to have a chance to achieve this potential – as we believe it does – there must first be a vision of what is possible. Topol has discussed at length the potential of accumulating more data; our focus here is on extracting more information from that data.

It is important to keep in mind that the medical domain has many stakeholders: patients, nurses, physicians, administrators and policy-makers, as well as clinics and hospitals. We believe that ML can make a revolution in healthcare – but it will not do so by replacing any of these stakeholders, but rather by enabling and empowering them to improve the entire path of healthcare – from prevention, to diagnosis, to prognosis, to treatment, with the ultimate aim of enabling "individualised medicine", while maintaining or even reducing costs. In particular, ML will support and complement, rather than substitute for, the judgment of medical personnel, and will also inform patients and administrators. Put differently: the purpose of ML in the medical domain is to provide intelligence – especially actionable intelligence – and decision support to all the stakeholders.

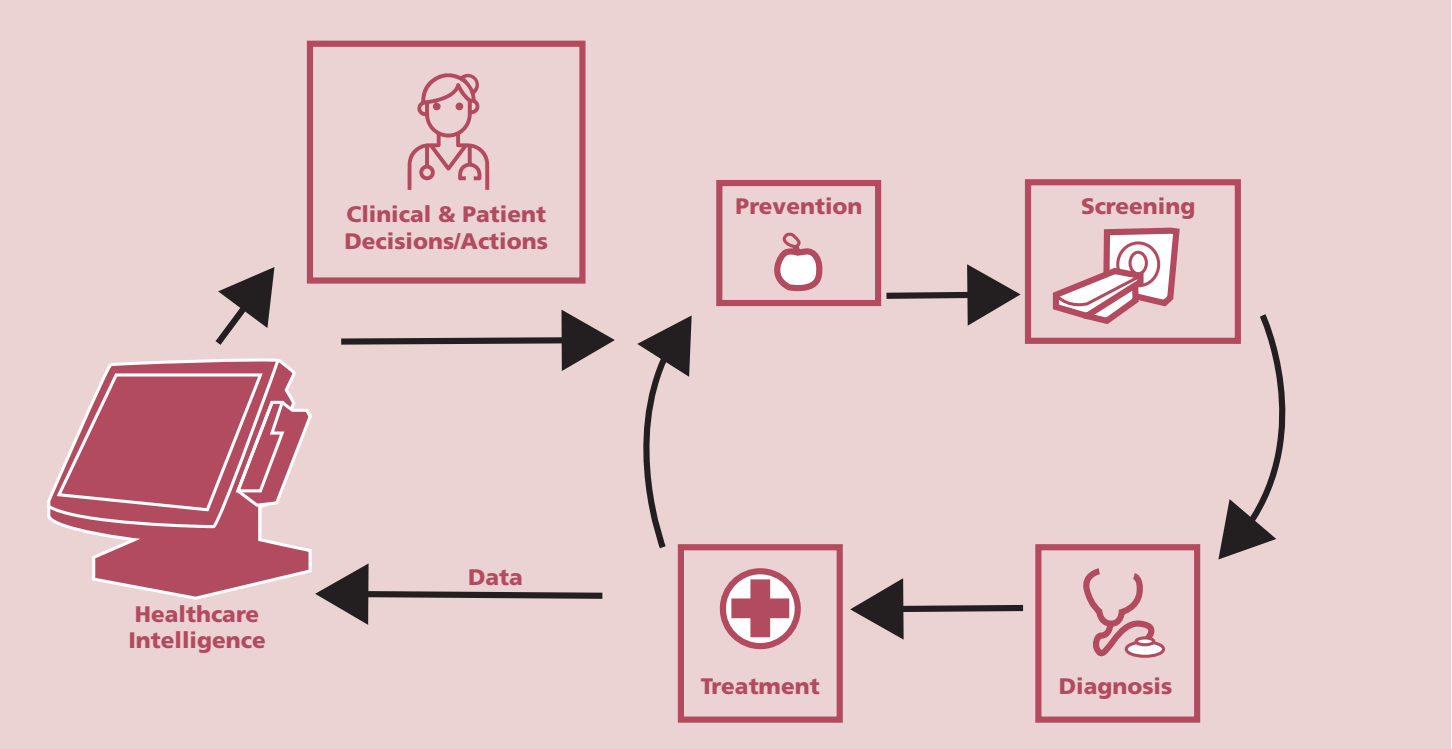
Currently a lot of the technology that has been developed by ML (and by digital medicine) has been driven by opportunities seen by the technological community, rather than by the need to provide information that matters, that is actionable, and that will actually improve healthcare. This must change: the development of ML for healthcare must be pulled by the healthcare stakeholders (from patients to policy-makers) and not just pushed by the technological community. The ML community and the medical community and stakeholders must work more as partners: the design and assembly of these building blocks must be guided by the needs of the users and supported by ML development. To accomplish this, ML must accomplish at least two things. The first is to provide building blocks – methods and algorithms – that the users can assemble for their own particular needs. The second is to make these methods and algorithms sufficiently transparent and understandable, and validated in a wide variety of contexts, that they earn the trust of the users.

It may be appropriate to begin by briefly discussing just one of the challenges that we believe ML can help to meet: the rising incidence and cost of cancer. Approximately 360,000 people in the United Kingdom are diagnosed with cancer each year and approximately 160,000 people die of the disease. The NHS currently spends approximately £6.7 billion per annum treating cancer patients. Moreover, as the average age of the population increases over the next three decades, the incidence of cancer is projected to rise – and the cost of treatment is projected to rise even faster. According to the statistics fact sheet issued by Macmillan Cancer Support[2], "there are an estimated 2.5 million people living with cancer in the UK in 2015, rising to 4 million by 2030. The number of people living with cancer in the UK in 2015 has increased by almost half a million people in the last five years." Given these numbers, it is clear that treating an increasing population of patients with current methods will rapidly become financially challenging. Similar problems exist globally: "the economic impact of cancer is significant and is increasing. The total annual economic cost of cancer in 2010 was estimated at approximately US$1.16 trillion".[3]

Although this challenge is daunting, it represents an opportunity for ML to provide better tools for the medical community to prevent, detect and treat cancer – and at a sustainable cost. Similar challenges and opportunities exist for many other diseases and throughout medicine. Especially important examples include the multitude of morbidities for which the aging population of the United Kingdom is increasingly at risk. It is to these challenges and opportunities that this chapter is addressed.

* We distinguish Machine Learning from the broader discipline of Artificial Intelligence, which refers to any system that can perform "intelligent" tasks – including systems that act entirely on the basis of pre-determined rules.

Figure 10.1 Machine learning in healthcare



Source Mihaela van der Schaar

Human beings are good at understanding and drawing conclusions from information gathered from a limited collection of data sources over a short period of time. They are less adept at understanding and drawing conclusions from information gathered from a broad collection of diverse data sources over a long period of time – although such information may in fact have significant impact on all aspects of decision-making. In the medical domain, this is particularly true because more and more data is becoming available from a wider and wider variety of sources (imaging, genomics, vital signs and lab tests collected over time, etc.). Extracting intelligence from this data is necessary to advance to new and improved levels of prognosis, diagnosis and treatment. This is especially important in dealing with complex diseases such as Alzheimer's, cancer, cardiovascular disease, cystic fibrosis, diabetes, etc. which exhibit complex phenotypes and require sophisticated analysis of data – many features, not just a few – which is difficult for humans but easier for machines (computers).

There is already too much data for human beings to keep track of and make use of. This is one of the reasons that clinical risk scoring methods tend to use only a subset of the features that are recorded: treating clinicians simply cannot keep track of and make use of the hundreds or even thousands of features that may actually be available. And the amount of data is growing exponentially quickly; not just the data in electronic health records (which is now recorded for a majority of patients in developed countries) but also genomic data, proteomic data, etc. Although some of this data (e.g. the genomic data) is fixed, some of it changes rapidly – and it is not just the current data that is important but also the history/trajectory of the data. Moreover, as guidelines are changing, different information is being collected. Static clinical methods have a hard time adjusting to these changes. ML methods can do much better because they can be continually re-trained to incorporate these changes and adapt to them.

Progress toward understanding, preventing and treating such diseases will require developing new methods that break the barriers between ML, statistics and mathematics, and using these new methods to discover new links, causes and causal relationships among clinical data, genetic data, metabolic data, environmental data, social data etc. and their implications on the risk, incidence and trajectory of disease.

In what follows, we discuss some of what ML has accomplished and can accomplish in a wide variety of areas. We begin by discussing risk scoring because it pervades every aspect of medical care.

02 Risk scoring

Evaluation of risk is essential in many medical domains because it informs every aspect of patient care, including prevention of a disease/condition and prognosis and treatment after the onset. Before the onset of a disease/condition, patients thought to be at high risk may be screened more often and more thoroughly, and are more likely to receive prophylactic interventions and/or treatment. After the onset of a disease/condition the course of treatment will often be different for patients having different levels of risk. Evaluation of risk also plays a particular role in operational settings where triage is key.

Currently, most evaluation of risk is not done systematically, or on the basis of any formal model, and hence does not integrate the wealth of information available about the patient. Even in areas in which the evaluation of risk is done on the basis of a formal model that produces specific risk scores, it is usually done on the basis of a linear model, such as a Cox proportional hazards model, and using a relatively small number of hand-picked features, often chosen either by clinical judgments or by simplistic variable selection methods. ML-based methods provide two kinds of gains over such models. The first is an informational gain: ML-based methods are able to handle many more features. This is especially important because different features often make different contributions to risk for different classes of patients, and it is becoming even more important as more kinds of data are becoming available – hence there are more features to be taken into account. The second is a modelling gain: ML-based methods are able to make better use of the same features by better capturing the potentially complex interactions between features. These gains allow ML-based methods to issue more accurate predictions, and hence better treatment guidance, for the patient at hand. Because ML methods provide both informational gains and modelling gains, the improvement in predictive accuracy of ML-based models over existing clinical models has been, and is likely to remain, largest for settings (diseases) in which many features are potentially important – especially if different features are important for different kinds of patients – and in which features interact in complicated ways that are not captured by linear models such as the Cox proportional hazards model.

An additional advantage of ML-based models is that – in part because they are capable of using many features – they make it possible to discover the importance of features and of interactions among features that were not previously understood to be important. For example, the work of Alaa and van der Schaar revealed an unexpectedly important role for oxygenation – in addition to FEV1 (forced expiratory volume) – in predicting the decline of patients suffering from Cystic Fibrosis.[4]

A number of challenges need to be overcome to enable large-scale deployment and use of ML-based risk-scoring methods. Some of these challenges, and the progress that is being made to meet the challenges, are outlined here.

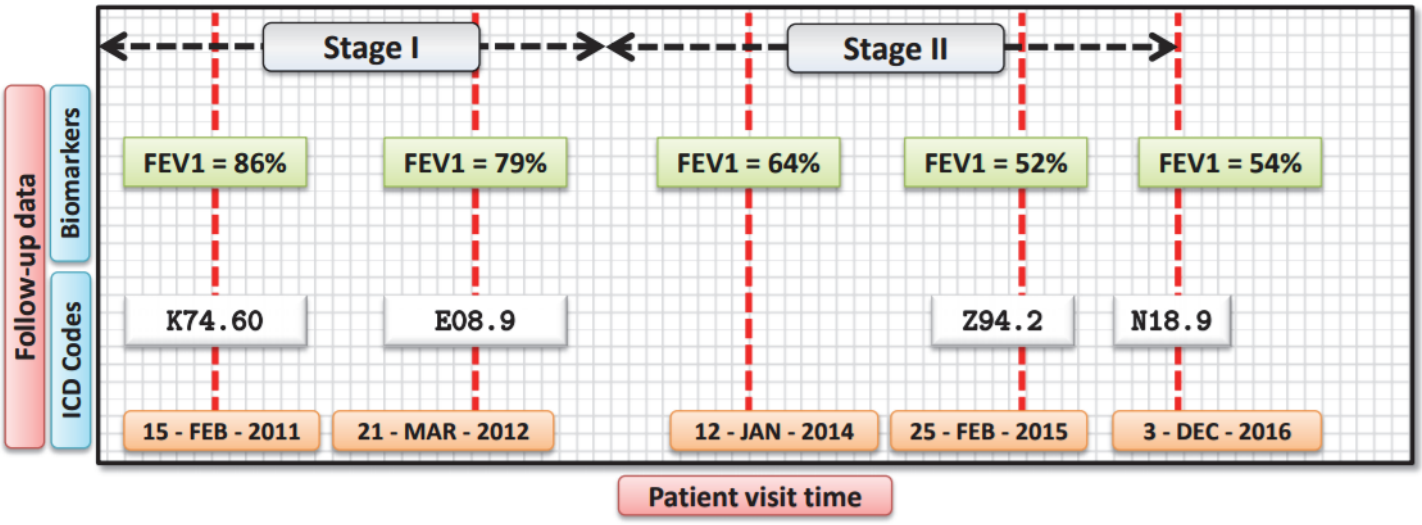
2.1 Longitudinal trajectories and multiple states/stages of disease

The most important challenge is the universal problem that there are multiple states/stages of conditions/diseases and the true state and the true probability of transition from one state to another cannot be observed directly. Instead it must be inferred from information that can be observed directly, such as symptoms, measurements, tests, images, etc. Figure 10.2 illustrates the problem: the patient transitioned from Stage I to Stage II – indicated by a large drop in FEV1 – between visits. Moreover, it is important to make use of what can be learned implicitly from the absence of information in the electronic health record of a patient. ML has made great progress in recent years on learning trajectories of risk and disease – and deriving risk and prognostic predictions – from the information available in the patient records.[5] ML has also made substantial progress in the development of new methods for understanding disease from longitudinal data, including the number of states needed to provide an accurate representation of the disease, how to infer the current state, what triggers the transition from one state to another, etc. – all on the basis of the available information. The progress that has already been made, and the progress that will be made in the future, can play a key role in understanding and preventing disease as well as in triaging patients.

2.2 ML Know-how

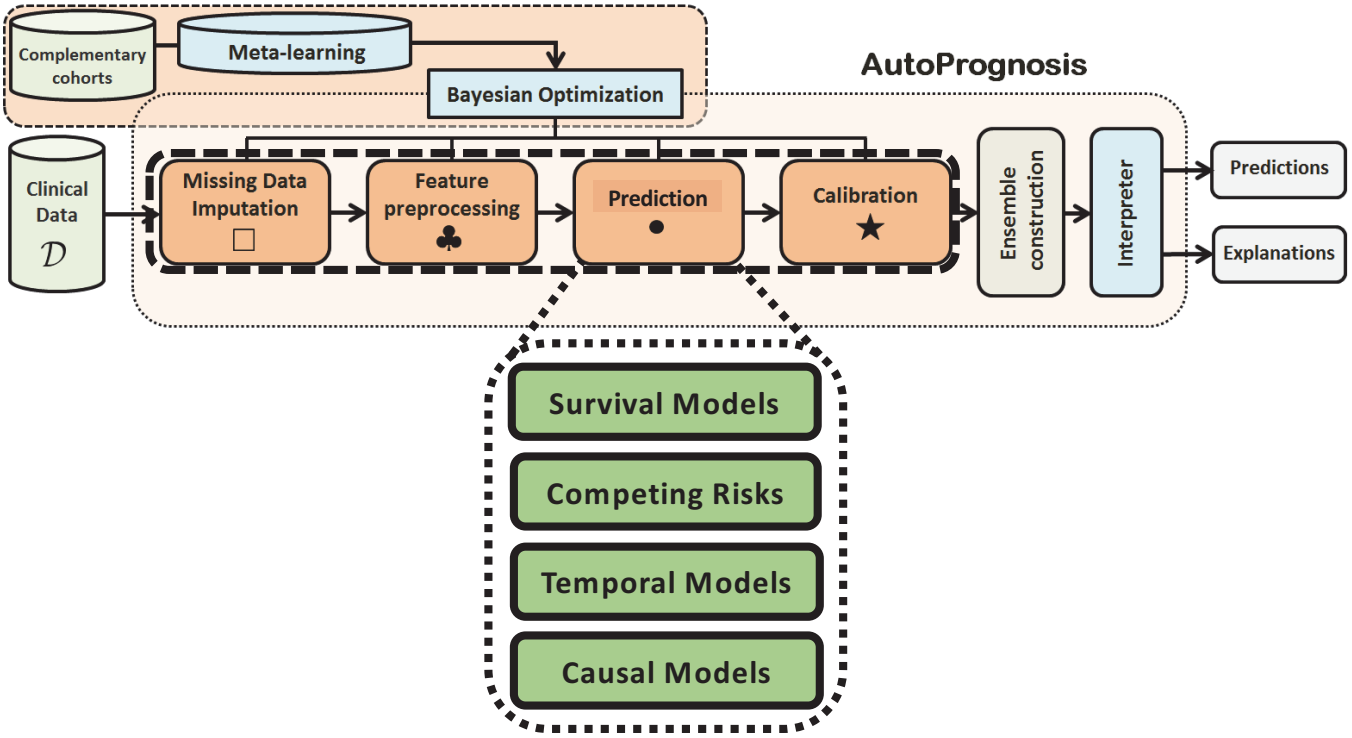
One of the barriers to the use of ML-based methods has been that they have required a great deal of technical knowledge of ML to choose and tune the particular predictive model to be used. However, the development of automated methods such as AutoPrognosis[6] is now making it possible for clinical researchers who have little or no technical knowledge of ML to apply state-of-the-art ML-methods. These automated methods use ML itself to both choose and tune the ML-model(s) – more generally, the entire pipelines of models – to be used. Figure 10.3 provides a diagram of such a pipeline.

Figure 10.2 An example of the risk and disease trajectory for a patient



Source Mihaela van der Schaar and Ahmed Alaa

Figure 10.3 AutoPrognosis – Automatic Machine Learning for Healthcare



Source *Ahmed Alaa, Mihaela van der Schaar, "AutoPrognosis: Automated Clinical Prognostic Modelling via Bayesian Optimization with Structured Kernel Learning," ICML, 2018*

2.3 Changing Practice

One of the challenges facing all predictive models is that medical practice is constantly changing – and often the population of patients is changing as well. For example, the management and care of patients who have had heart failure and are potential candidates for a heart transplant has changed dramatically over the past decades as surgical procedures and assistive technologies have improved. ML-based predictive models such as the one developed by Yoon, Zame and van der Schaar can more easily adapt to these changes than traditional clinical models, both because they can incorporate more features and because they are more readily re-trained to incorporate new data.[7,8]

2.4 Interpretability

In order to be integrated into clinical practice in truly useful way, it is not enough that models be accurate – they must also be understandable to the users. This requires that the models have a high degree of transparency and that the predictions of the models should be interpretable to the users, who will be accountable for the decisions made on the basis of these predictions. Many widely-used clinical models have these properties – but many ML models do not. Rather they are "black-boxes" whose workings are rather opaque and whose predictions are not readily interpretable. This has posed a barrier to clinical acceptance of ML models – even when the data has shown that ML models provide more accurate predictions.

Interpretability is important for the adoption of ML methods in clinical practice. In order to be adopted, it is not sufficient that a new model predicts better than an older model; it is also necessary to understand why. Does the new model make better use of features that are already known to have clinical relevance (e.g., by better capturing the interactions between features)? Does the new model make use of features that were not previously known to have clinical relevance, but whose clinical relevance can now be understood? Does the new model identify subpopulations for whom certain features – or the interactions between certain features – play a different role than in the general population? The absence of satisfactory answers to these and other similar questions may lead to dismissing a new model on the grounds that its predictive success is simply coincidental (e.g. specific to a particular dataset).

There are (at least) three approaches to the challenge of interpretability. The first is to focus on refining existing ML models that are (relatively) transparent and whose predictions are already interpretable. Classification and Regression Trees have these properties. The second is to create new models with these same properties; this is part of the motivation for the creation of a new generation of ML models such as Trees of Predictors.[7,8] The third is to provide methods for interpreting existing ML models; this has prompted a substantial recent body of work.[9,10]

An important aspect of all of this is that much recent work has pointed to a trade-off between transparency/interpretability and accuracy. This is a trade-off that the ML community can identify – but it is up to the medical community to decide how to make the trade-off. Clinicians who will be accountable for the decisions made on the basis the model they use might (quite reasonably) emphasize transparency and ease of interpretation in their choice of which model to use. Thus, it is important that the users of the models be involved in the process of creation and evaluation.

2.5 Leveraging Multiple Datasets

Modern ML methods often require large amounts of data in order to learn the large number of parameters that define them efficiently; this may be because the model itself is complex and/or because the dimensions of the input data are large. Acquiring large datasets can be difficult. Moreover, it is often important to learn a model that performs well on a specific, potentially small, sub-population. Learning such models can be difficult because data from a particular hospital is often limited, so it is important to leverage data from other hospitals, while keeping in mind that different hospitals may collect different data, may serve different populations with different incidences of disease, and may treat patients differently. Hence, simply importing data from other hospitals may therefore lead to biased and inaccurate prediction. ML-based methods are capable of learning to translate from one clinical setting (dataset) to another, thereby effectively enlarging the dataset from a single hospital and enabling the creation of better predictive models.[11]

2.6 Multi morbidities

Some patients suffer from – or are at risk for – multiple diseases/conditions; these risks significantly increase as the patient ages. In order to monitor and treat such patients, it is important to predict which disease/condition is likely to occur sooner and which is likely to occur later and how the risks for various diseases/conditions are changing over time. This poses problems for models that consider each risk separately (e.g the cause-specific Cox model) and often leads these models to make very poor predictions; what is necessary is to treat the multiple competing risks in a holistic fashion. ML-based methods have done much better by learning representations of factors that are shared between the diseases/conditions and hence produce joint predictions of the risks of the various diseases/conditions.[12]

Most of the work, however, that has been done on multiple morbidities has focused on the setting in which patients might develop any of several diseases, but will actually develop only one. For example: a patient might be at risk of death from both cancer and cardiovascular disease – but will die only of one. However, it is often the case that a patient may develop first one condition – e.g. diabetes – that does not result in death but that affects the risk of developing a second condition – e.g. cardiovascular disease.

This latter setting is especially important because more preventive care is potentially possible. Initial work in this area[12,13,14] suggests that ML may have enormous promise in dealing with this problem.

As should be clear from this discussion, ML has enormous potential in the domain of risk-scoring. It is important to keep in mind that the potential of ML in the domain of risk-scoring is not only informing medical personnel but also informing patients themselves. Individualised information about risk factors, current and predicted health state, preventive measures and treatment plans can be provided to patients and embedded into apps and wearable devices, which can track the health state and risks of the patient over time, and empower patients to make informed healthy choices, adopt better lifestyles and better adhere to plans for screening, monitoring and treatment so that disease is prevented or treated effectively at early stages. This can lead to a new equilibrium in which patients are more involved in their own healthcare.

03 Imaging and diagnosis

Medical imaging and diagnosis is perhaps the area in which ML has had the greatest success so far.[15] Some of this success (especially the success that has come from deep learning) is very widely known, and has been trumpeted in the popular press. ML has demonstrated its ability to provide more accurate and definitive interpretation of images, especially in the areas of radiology, ophthalmology[16,*], pathology[17-19] and dermatology.[20,21] These methods can be incorporated into decision support systems that can aid physicians by offering second opinions, flagging concerning areas in images and determining abnormal tissue or lesions. The progress in ML in these areas has exploited the enormous body of existing data (millions of photographic images), the ability to easily generate additional data, and the exploitation of a great deal of prior knowledge and understanding of the structure of the data.

Although these achievements are impressive, much more is possible. A particularly promising new application of ML to medical imaging is in digital mammography. Improved accuracy of digital mammography could have a profound impact on breast cancer screening programs by reducing manpower requirements, reducing costs and distinguishing patients who are at greater risk from patients who are at lesser risk. This need for better risk prediction is already recognised and steps toward an adaptive approach to screening are already being undertaken. For example, women with "dense breasts" are often directed to supplemental screening modalities such as ultrasound, contrast enhanced and/or 3-dimensional mammography or MRI scanning. However, "dense breasts" is a very coarse – and often subjective – classification; ML can more objectively identify which of these expensive and resource-intensive technologies should be used for which women – and when.

ML can also play an important role in the interpretation of ordinary mammograms. Properly trained and validated ML algorithms could be used to assist radiologists in a number of ways. Most obviously, ML algorithms can identify some images that are not problematical and do not need to be examined further, thus providing the radiologist with more time to examine those images that are problematical. Perhaps less obviously, ML algorithms work with the radiologist in examining images that are problematical – e.g. identifying areas that appear suspicious. The ultimate aim is to use ML to streamline the reading process, allowing radiologists to focus on more difficult cases and improve diagnostic accuracy. By answering and understanding what ML methods – especially deep learning methods – are good at, we can take the next step to integrating them into an operational clinical setting by instilling confidence through interpretability and understanding.

More generally, ML can improve breast cancer screening and detection by better stratifying patients into risk groups, and determining which groups would benefit most from different screening regimes. These various regimes might be differentiated by the time at which screening begins, the frequency of screening, the particular imaging modality/ modalities (mammogram, 3-D mammogram, ultrasound, MRI with/without contrast, etc.) to be used at each screening opportunity, etc. The initial assignment of a patient to a particular risk group (and hence to a particular screening regime) should be based on the attributes of the patient, including demographics, lifestyle, family history and genetic information, but the assignment should change as new information is provided by screening that has already taken place. The information provided by this more individualised screening process will help to provide support for clinicians in deciding when a biopsy is needed, and perhaps what steps to take after the results of a biopsy are known. See Figure 10.4 for an overview of what the process could look like, and for some steps along the way.[22]

More generally, leveraging digital pathology and advanced radiological imaging and integrating these with ML can increase the efficiency, speed and accuracy of detection, diagnosis and monitoring for many diseases.

* See Chapter 12 of this report for a case study of AI in ophthalmology (Moorfields-DeepMind Collaboration, 2016-2018)

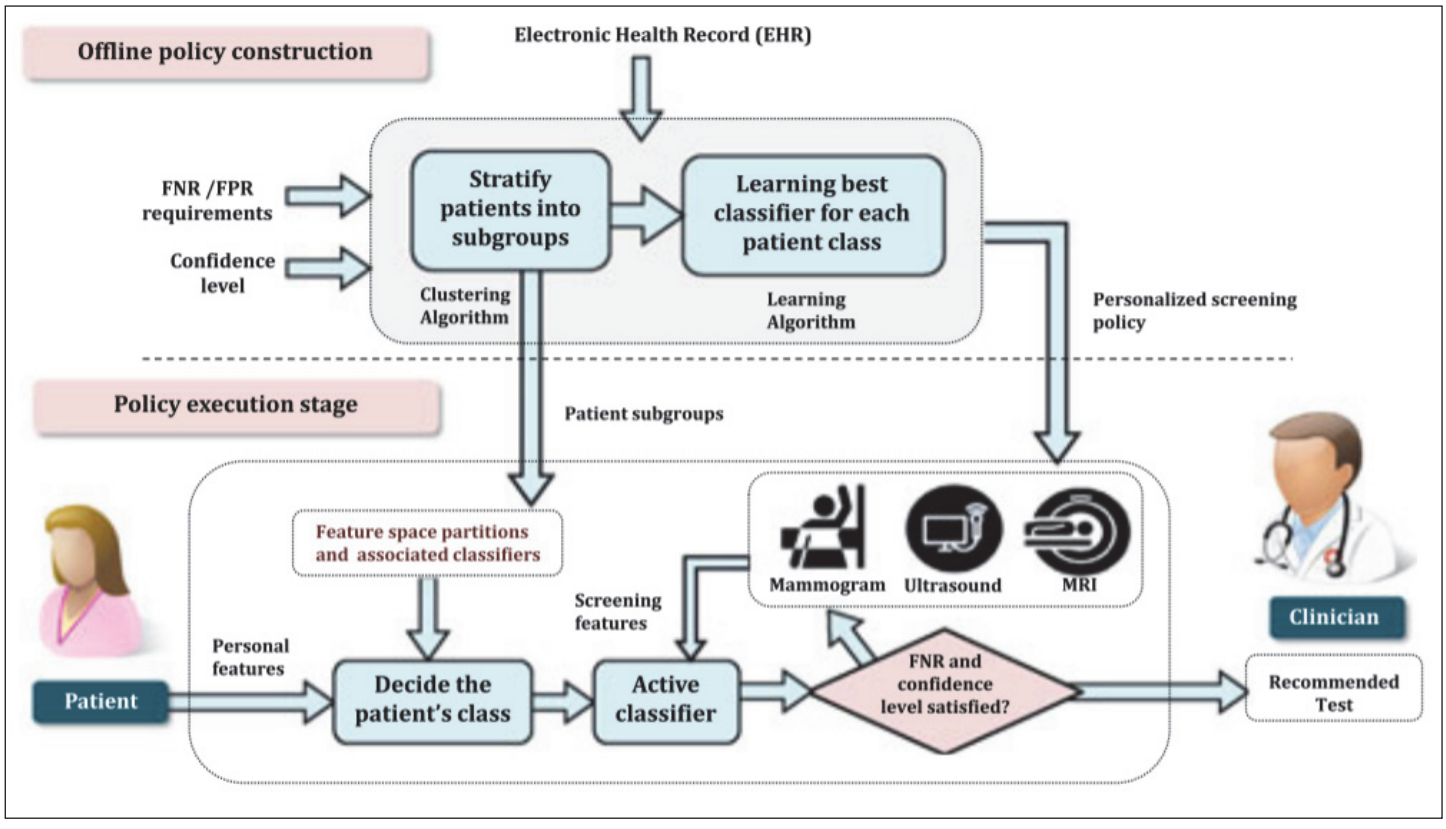
04 Screening

For most diseases, the current guidelines prescribe a one-size-fits-all screening program: a particular screening procedure (which may involve imaging, but also many other tests) is carried out at fixed intervals for every patient in the relevant population. However, for many diseases, the actual risk faced by a patient varies widely across patients, and for the same patient, may vary widely over time. Hence patients whose actual risk is low are screened too often – which is wasteful of societal resources and unnecessarily costly/invasive for the patient – while patients whose actual risk is high are not screened often enough, so that detection of disease is delayed. It is therefore important to tailor screening policies to the particular patient.

For a few diseases, such as breast and colon cancer, current United Kingdom screening policies do take into account genetic markers for the disease, family and personal history of the disease, and the findings of previous screenings. However, even for such diseases, current policies group patients into very large bins and follow a single procedure within each bin. We have already mentioned that ML has the potential to greatly improve digital mammography, but this is only a part of its potential to improve breast cancer screening. ML also has the potential to integrate a wide variety of patient features with the patient's history, to create individualised screening policies that are better for the patient while also making better use of expensive and scarce resources. Moreover, even after initial screening has revealed the presence of a condition/disease, treatment and follow-up are typically either one-size-fits-all or very ad-hoc – depending on the experience and judgment of the treating clinician and perhaps the insistence of the patient – despite the fact that actual risk varies widely across patients.

It is important to keep in mind that the recommendations of such a policy will sometimes be "screen more often" or "use more accurate – but perhaps more invasive or more expensive – screening procedures". They will also sometimes be exactly the opposite, because many patients who are low risk do not need to be screened as often and do not need costly or invasive screening procedures. Less frequent screening of patients at low risk frees up scarce resources to provide more frequent screening for patients at high risk. In addition, ML has the potential to learn the value of screening and of screening over time for the particular patient at hand. And it is not only the timing of screening that should be individualised to a particular patient: ML can learn the value of information – and the value of information over time – and hence learn that some patients require a particular method of screening or follow-up while others do not.

Figure 10.4 Building an individualised screening policy



Source Ahmed Alaa, Kenneth Moon, William Hsu and Mihaela van der Schaar, 2016

05 Individualised treatment effects

Choosing the best treatment for a particular patient requires estimating the effects of a variety of treatments for that particular patient – not only some average over the large population of patients. To accomplish this, we need to understand individualised treatment effects and not just average treatment effects. Unfortunately, this is beyond the state of the art of current medical knowledge.

Clinical trials are – and will remain – the gold standard for understanding the efficacy of new medications and treatments. However clinical trials, by their very nature, have important limitations. They have strict criteria for entrance that often exclude elderly patients and patients with multiple morbidities (which often represent a large fraction of the patients who will actually receive the medication or treatment), and they are often limited by ethical considerations. Moreover, the guidelines developed on the basis of clinical trials are based on the average effects found in the trial, although it is often the case that a given medication or treatment may have very different effects for some particular subpopulations than for the population as a whole. Clinical trials often provide a lot of information about average treatment effects but much less about individualised treatment effects.

Recent advances in ML have shown that ML is capable of learning individualised treatment effects by beginning where clinical trials leave off – learning from observational data: the data captured in clinical practice. Learning individualised treatment effects from observational data is an extremely challenging problem because the data is biased (clinicians choose which patients should receive the drug/treatment rather than randomising), counterfactuals (what the outcome would have been if a treated patient had not been treated or conversely if a non-treated patient had been treated) are not observed, and both the decisions and the outcomes may be affected by hidden confounders (not recorded in the observational data). Moreover, because counterfactuals are not observed, predictive models cannot be tested out of sample.

However, recent work in ML has made promising progress on this problem. Stefan Wager and Susan Athey have adapted a standard ML-method (Random Forest) in order to produce predictions that come with statistically valid confidence intervals.[23] Ahmed Alaa and Mihaela van der Schaar have developed general guidelines for the design of ML algorithms for estimating individualised treatment effects by characterising the fundamental limits of what can be achieved and establishing conditions under which these fundamental limits can be realised.[24] Their theoretical analysis reveals, among other things, that the relative importance of the different aspects of observational data vary with the sample size. Selection bias is the most important bottleneck to performance in small-sample regimes, while the modelling of the control and treated outcomes is the most important bottleneck to performance in large-sample regimes. Building on these findings – and others that result from the theoretical analysis[24,25] – they develop practical ML-algorithms that outperform previous methods.

The impact of accurate prediction of individualised treatment effects is hard to overestimate because it provides a basis for the choice of treatment according to the features of the particular patient at hand. This leads to improved outcomes for patients – who receive the best treatment – and for improved use of resources, not wasting scarce resources on patients who will receive little or no additional benefit from them. One potent example comes from the realm of cardiac transplantation: how much would a patient who has experienced heart failure benefit from a mechanical assist device (e.g., an LVAD), either as a bridge therapy or as a destination therapy? A second example comes from the realm of chemotherapy: when one regimen of chemotherapy has achieved all that it can, what – if any – further regimen should be tried next?

While ML methods suggest that substantial progress has been made in understanding individualised treatment effects on the basis of observational data, the progress so far has serious limitations. Most importantly, the findings of ML with regard to individualised treatment effects need to be validated in clinical practice. One obvious way to do this is to use the predictions of ML to design further clinical trials, perhaps focusing on subpopulations for which the effect of a particular treatment seems different from the effect on the population as a whole.

06 Monitoring and early warning systems

Many hospitalised patients suffer sudden, unexpected and life-threatening deterioration such as septic shock or cardiac arrest, and the death rate from these events is very high (as much as 75%). However, many of these events could have been prevented, and many other patients who did experience these events could have been saved, if earlier warning – even a few hours – had been given. Unfortunately, the existing early warning systems that are widely used in hospitals have not proved adequate: they miss many events and issue many false alarms – which strains the resources of hospitals and leads to alarm fatigue among care-givers.

Providing accurate early warnings – both in and out of the hospital – is a challenging problem for many reasons, of which we mention only a few. The most important reason is the universal problem that the true state of the patient cannot be observed directly but must be inferred from what can be observed directly. Such inference requires integrating data of many kinds from many sources: demographic information, laboratory tests, vital signs, imaging, etc. Moreover, since observations cannot be made constantly, the possibility that the patient's state has changed in between observations must be taken into account. And the probability that a patient will deteriorate in the near future depends not only on the patient's current state but also on the past history – in particular, on the time the patient has been in the current state.

Despite these difficulties, ML methods have made real progress. Saria et al have developed ML methods that provide predictions of septic shock that are more accurate and more timely than those provided by commonly-used clinical models.[26] Van der Schaar et al have developed ML methods that provide predictions of cardiac arrest in hospital that are more accurate and more timely than those provided by commonly-used clinical models.[27-29] The methods used in these two examples are very different, in part because they address very different kinds of sudden deterioration in hospital. But there are many kinds of sudden deterioration in hospital, and creating a predictive model for each one would present an insuperable task. Here again, the idea of Auto-Prognosis[6] – to make ML itself create the model – has enormous potential.

Outside the hospital, a variety of sensors of many different types, including wearables, are becoming widely available for use by consumers. These sensors can already track activities, motions, food intake, body temperature, heart rate, blood pressure; in the very near future they will also be able to track blood chemistry and many other things. These sensors can enhance care in (at least) three ways:

- By collecting information that can be used by the treating clinicians (especially in tracking patients' behaviour and progress).
- By collecting information that can be used by the patient to monitor, track, follow progress.
- By providing feedback to the patients including warnings, reminders and suggestions for adherence to medication, exercise and treatment regimes.

ML will play a vital role in this by creating systems that can: integrate the data; make predictions about the patients' health, behaviour, adherence to medications and therapy, etc.; provide feedback about the patients' progress; offer recommendations including alternative regimes of diet and exercise; and alert the patient and medical personnel when further consultation is needed. Such systems can also empower and encourage patients to take a more active role in their own healthcare by providing them with feedback about the (positive or negative) effects of their own actions and behaviours. They can also increase the efficiency of the patient/clinician relationship by keeping both informed about the necessary level of interaction. Is a visit necessary? Should the physician provide feedback about progress (or the lack of it)? Should the physician suggest changes in behaviour? in medication? in treatment?

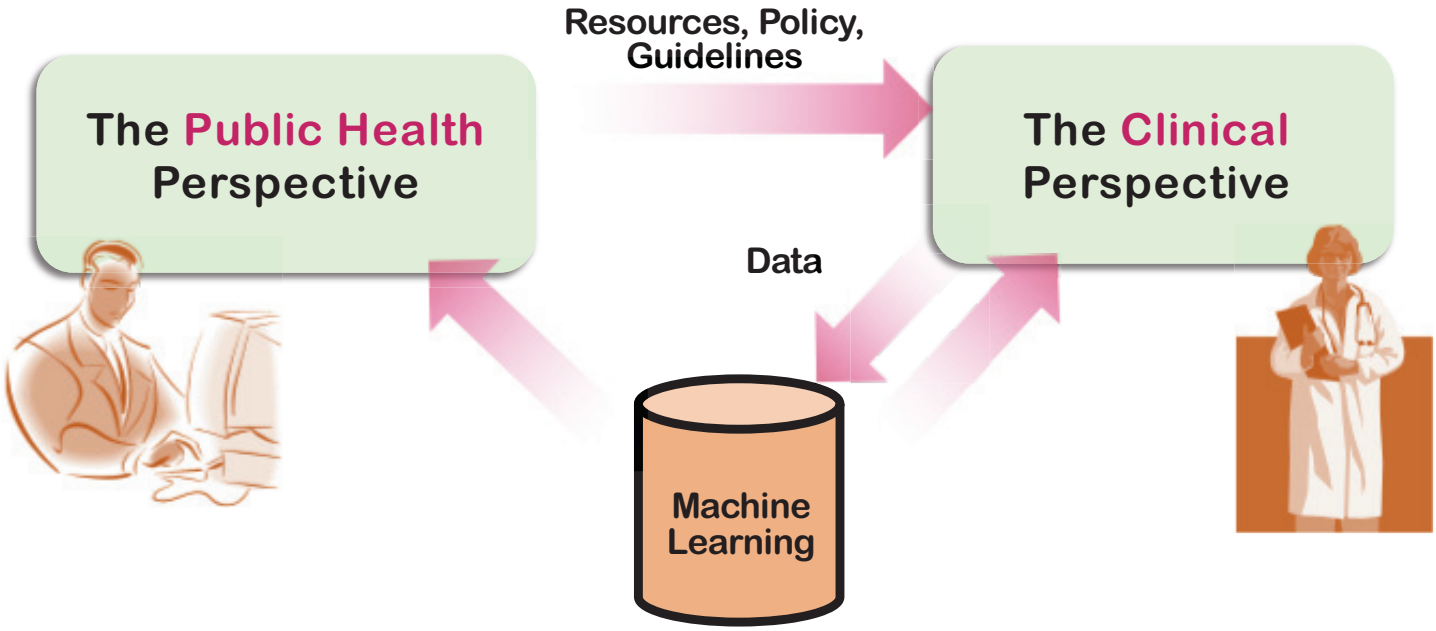
An important aspect of the feedback loop that remote monitoring (especially wearables) enables is that the clinician may become more involved with the patient – even though the patient may make many fewer visits to the clinician.

07 The electronic health records of the future

We have discussed many ways in which ML can improve health care, but one of the most important ways is by creating a system that integrates all of these. This might be thought of as the electronic health records of the future – but since it will not only record data but also process and integrate information and provide forecasts and recommendations, it might better be thought of as a Learning Engine for Healthcare (LEH).* In addition to all the information currently recorded in electronic health records, the LEH will produce a holistic view of risk – and the trajectory of risk – for many diseases to which the current patient might be at risk, and select and display those that are requested (by the clinician or patient) and those that are most pertinent, even if not specifically requested. This should be contrasted with current electronic health records, which can compute only risks for a few diseases, and display only those which are specifically requested. The distinction is important because many patients may actually be at risk for diseases that are not suspected (by the clinician or patient). By anticipating risks, the LEH becomes a tool by which the clinician and patient can actively prevent future disease (through lifestyle modification, prophylactic treatment, etc.).

Because the LEH can continuously learn from observational data that is recorded in the EHR and information that is provided by patient wearables or apps, the LEH can continually update the intelligence provided to the user, leading to a cycle of improvement in all stages of care – from prevention to screening to diagnosis to treatment.

Figure 10.5 ML informs both the public health perspective and the clinical perspective



Source Mihaela van der Schaar and Ahmed Alaa

* See Oon International Award and Lecture in Preventative Medicine 2018 for some more details about the Learning Engines for Healthcare and their capabilities (https://www.youtube.com/watch?v=2MdvYoN6_20)

08 Public health

With the availability of more diverse data, such as the data that is now becoming available in UK Biobank, wearables, social media and a plethora of apps that track exercise, nutrition and vital signs, ML can unravel a variety of determinants of health; not only clinical factors but also social, environmental, nutritional and behavioural factors. This provides the potential to identify and to personalise the determinants of health (risk factors) that go beyond well-known clinical markers. (For example, it is already well-known that high cholesterol is generally associated with higher risk of cardiovascular disease, but ML methods are being developed that can assess the extent of this risk and suggest appropriate counter-measures – which may be very different for different individuals.) To do this, there are many challenges that need to be addressed and which ML has been shown capable of addressing. The first is that the number of risk factors – which now include environmental, social and behaviour factors – is vast, and hence selecting those relatively few factors that really matter and how those factors interact requires new methodology. The second is that different factors matter for different people so it is essential to discover those factors that matter for each individual. The third is that different factors matter at different times, both at different ages and at different stages of disease. The fourth is that the factors that matter for a given disease may be different when the patient suffers from – or is at risk for – another disease. Thus, a holistic view is a necessity. ML provides the methodology to address all these challenges.

This can have significant impact on public health, by enabling a shift from providing a one-size-fits-all view of one disease at a time to a more personalised and holistic view.

09 Communication and interaction

It has been estimated that as much as 50% or more of physicians' time is spent simply entering data.[30] The data obtained in this way is invaluable, and its accessibility in electronic health records makes it extremely usable. However, because the data is collected by human beings, the collection of the data and its entry into the electronic health record are prone to error and very time-consuming. Natural language processing (NLP) has enormous potential to reduce the time for data entry and the probability of error. (Natural language processing (NLP) is a subfield of AI and ML concerned with the programming of computers to process, analyse, understand and interpret language data. Challenges in NLP involve speech recognition, language understanding and interpretation, and generation of language for clinical notes and reports, etc.[31]).This has the potential to be transformative because it will reduce the time spend on data entry and free up valuable time for actual interaction with patients. It also has the potential to increase the amount of data that is routinely collected and entered into the record while reducing the number of entry errors.

A second and perhaps less obvious use of NLP is in the analysis of text of many kinds, including clinical notes, imaging reports, etc., and in integrating this text with other sources of information as an additional set of patient features, or an additional set of labels, or as a way to enable semi-supervised learning. Raw text is difficult for ML methods to use, but the additional information in this text may prove extremely important in facilitating and enhancing the performance of ML-based methods in order to improve prognosis, diagnosis and treatment.

A third application of NLP is in the development of chatbots, which may prove useful in a variety of settings, including preliminary triaging in NHS 111 services and patient support of all kinds including mental health, adherence to medication therapy and treatment, reminders, etc.

Another important aspect of communication is the contextual interaction taking place among the participants in the treatment and support of each particular patient. The treatment of complex diseases such as cancer typically involves many clinicians across many sub-specialties in addition to the GP, the patient support network and the patient. Rapid and accurate communication among all these participants is essential for the best treatment. This is yet another area in which ML can be of enormous value by automatically creating networks of expertise and support, based on the state of the patient, that will identify information that needs to be solicited and shared and by providing the means in which that information can be shared in ways that are easily understood – and, when appropriate – acted on by the various participants. (For an example of this see Tekin, Atan and van der Schaar.[32]) This may require the integration of many areas of ML including multi-agent learning, contextual learning, natural language processing and others.

10 Operations management in healthcare

It is currently estimated that a very large fraction – perhaps as much as 50% – of US healthcare spending constitutes waste, and that this waste largely arises in a few categories: failures of care coordination, failures in execution of care processes, administrative complexities, pricing failures, overtreatment, and fraud/abuse. ML can significantly reduce this waste by streamlining the operation of the entire system of providing healthcare, thereby reducing costs and improving outcomes. For example, in the domain of care coordination, ML can advise medical personnel, staff, administrators and patients with whom an appointment should be made (nurse, general practitioner, specialist) for a specific patient with a specific history and current symptoms, when this appointment should occur, how much time should be scheduled for this appointment, and where this appointment should occur – in a clinic or remotely by telephone or videoconference. ML can also be tremendously useful in identifying and reducing fraud/abuse. (It has already been used to identify credit card fraud).[33]

Another way in which ML can also be enormously useful is in streamlining the delivery of care in the complex ecosystem of hospitals, for example by improving the triaging of patients at many stages. In the emergency department of a hospital, ML can help personnel to decide in which order patients need to be seen/treated: which patients must be seen/treated immediately and which patients can wait – and for how long – and which tests need to be performed to make these triaging decisions. ML can also play a crucial role in estimating the current disease state of the patient and determining when a patient requires immediate hospitalisation and surgery or can wait safely at home and be monitored remotely, thereby reducing the stress on the system and on the patient.

11 Education

ML can play an important role in training both medical students and practicing clinicians in a personalised way, in both traditional clinical areas and in non-traditional areas such as statistics and data science, including ML. Personalised education/training is especially important for practicing clinicians, who must constantly upgrade their knowledge and skills, but have no time to attend formal classes and may have very individual needs and different backgrounds/skills. Personalised education/training will accelerate the ability of clinicians to absorb new technology, the ability to re-train in new areas and especially the ability of clinicians to think more holistically about the patient in front of them. ML can help clinicians to learn about new medical advances and techniques and also about new ML-enabled decision support systems (as described above). ML-enabled personalised education is an active area of research[34] but personalisation to the special requirements of clinical and medical personnel poses special problems and will require additional work.[35]

12 The way forward

As we have noted in the beginning, we believe that ML can make a revolution in healthcare – but by empowering medical personnel (clinicians and researchers) rather than replacing them. Hence medical personnel should embrace what ML can do – rather than fear it – and should fully participate in – even lead/direct – the development of ML for healthcare. This is a positive message – and also a call to action for medical personnel, especially young personnel, to join – indeed drive – the revolution.

One of the reasons that medical personnel must be actively involved in this revolution is that the application of ML to medicine will require a different approach than has been taken in the application of ML in other areas. ML has been enormously successful in domains – such as game-playing[36] and autonomous vehicles – in which we have good models, in which the interactions (among players/entities and with the environment) are governed by "rules" (of the game, of physics, or of law) that can be clearly specified, in which data is plentiful or can be readily generated, and in which, whenever a decision – a particular move or sequence of moves, for instance – has led to failure/loss, it has always been possible to go back and try a different decision to see if the alternative would have done better.

Medicine presents a much more complex domain because we do not have good models, the interactions are governed by "rules" that we do not always understand – indeed, discovering the "rules" that govern disease and the human body is perhaps the most difficult task – and data is limited and fixed – we must learn from the patients we actually see rather than (hypothetical) patients we would like to see in order to gain more information. Moreover, in medicine, it is not possible to go back and undo a particular decision (surgery or treatment); in medicine we do not have, and cannot generate, counterfactuals. Despite this, ML for medicine has made substantial progress – and it promises to make much more.

Of course, the application of ML to medicine is really in its adolescence. New models need to be developed and validated, and their impact on efficacy, safety, quality and cost of healthcare needs to be properly assessed. Moreover, the opacity and complexity of many ML algorithms presents a barrier to their adoption. Widespread adoption of ML methods will require better understanding of existing algorithms and development of new algorithms that are more transparent and more easily interpreted. And throughout, attention must be paid to the central theme that ML algorithms and recommendations must go hand-in-hand with human judgment and actions; the combination of humans and machines will prove much more powerful than either alone. This is an extensive and ambitious agenda – but an important one.

We believe that the NHS provides a particularly fertile environment for the development and integration of ML into healthcare because it is a single party provider with a unified healthcare network that can promote rapid dissemination of innovations to an entire population. Close working relationships among NHS Trusts and allied academic institutions, provides the United Kingdom with the opportunity to lead the world in the transformation of healthcare through ML.

13 Authors' suggestions for policy

- Electronic Health Records are currently used only for the collection and storage of data – but their potential value as a resource goes far beyond this; they should be made more open and accessible. This will enable the extraction of actionable intelligence on behalf of the patients.
- Risk and treatment should move away from the current view that looks at a single risk at one moment in time to a view that looks at multiple competing risks over time. This will enable a holistic view of patients and patient care.
- Screening/monitoring policies should be personalised to the patient at hand – rather than based on one-size-fits-all clinical guidelines. This will lead to better use of resources, improved outcomes for patients and reduced costs.
- Learning individualised treatment effects on the basis of observational data should be used to complement learning on the basis of clinical trials. This will enable better matching of drugs/treatments to specific patients.
- ML should be integrated with Operations Research to improve triaging, queuing and work-flow planning. This will lead to better use of resources, improved outcomes for patients and reduced cost.

Box 10.1 A fictional narrative of big data and cardiometabolic disease personal prevention in 2040

Text kindly supplied by Mihaela van der Schaar, University of Cambridge and Jonathan Pearson-Stuttard, Imperial College London

Jim is 52 years old and lives with his wife and 2 teenage children in Newcastle upon-Tyne, England. Jim rises early ahead of a big day at work. Years ago, he used to pick up breakfast on his way to work, but now he doesn't need to. Jim upgraded his fridge to an 'intelligent fridge'; it re-stocks automatically, and does so according to Jim and his family's preferences, and the national guidelines for healthy dietary patterns.

Jim noticed his yoghurt was tasty, but different, this morning. He remembers his Electronic Health Engine (EHE) told him yesterday that his blood sugar is a little high, classed as 'pre-diabetes', and his food shop was modified by his artificial intelligence (AI) dietician and intelligent fridge. Jim enquires about the yoghurt – "Yes, this yoghurt has less sugar", Jim's smart speaker informs him.

Jim leaves home and walks to the bus stop; he is reminded by his (EHE) that this helps keep him healthy, but most importantly, helps him stay fit enough to keep up with his daughter, Jennie, when they play football together on Saturdays.

As Jim boards the bus, his (EHE) reminds him he's got a video consultation with a doctor once he arrives at work, in 20 minutes. Jim's (EHE) integrates all health and health-related data about Jim including social demographic data, food and alcohol consumption, physical activity, clinical measurements and his genome. Jim's (EHE) noticed his blood pressure, measured by his wearable health device, had been high consistently for past few weeks so it took 24-hour ambulatory measurements yesterday ahead of his video consultation today.

Jim's (EHE) scans all food and drink barcodes to update his risk trajectory and provide personal advice. This morning, his (EHE) advised against the packet of crisps Jim was going to buy this morning, offering lower-salt alternatives that are 'better for Jim'.

Jim has his video consultation before beginning his shift. He understands his recent dietary changes have helped a little and he should continue with them, but he would also benefit from beginning an anti-hypertensive drug. Jim's (EHE) information is available to his clinician via cloud computing, providing information regarding the most effective drug and dosage for Jim given his genome. Jim and his doctor agree to have another video call in one week to review Jim's blood pressure and new medication.

NB Throughout this scenario, Jim and his family have given their permission for data about them to be used in selected circumstances.

Box 10.2 Vision for Machine Learning in Cystic Fibrosis

Text kindly supplied by Oli Raynor and Mihaela van der Schaar

In Cystic Fibrosis (CF), patients have to continuously make life-changing decisions, and sometimes 'life and death' decisions, on the basis of data and evidence about the 'average' person with CF. But we know the 'average' person with CF does not exist. The disease affects people in very different ways. So, population-based evidence is of limited practical value to individuals trying to make medical decisions, manage and plan their lives, and understand their future.

At the same time, people with CF face a gruelling daily burden of self-administered treatment (typically 2-3 hours a day of pills, nebulisers, inhalers, injections, airway clearance and exercise) and there are no days off. There have been major advances in treatment but each new treatment adds to the existing regimen. It is hard to know whether a particular treatment is necessary for an individual at a given time. Are new treatments helping? Are old treatments still necessary?

People with CF also have to go to the clinic every three months for routine appointments, even when well. Cross-infection risks between people with CF mean they are advised not to meet in person – the time when it is hardest to follow this advice is when they attend hospital. This model of care is based on a one size fits all approach.

Machine learning has potential to enable a more personalised model of care based on more individualised characteristics and predictions of health trajectory. Decisions about whether a clinic visit is necessary could be informed by patient-generated health data gathered and recorded on digital devices. In fact, machine learning approaches might even suggest when an individual is about to have an exacerbation and proactively suggest some clinical action. When a new treatment is started, data helps people with CF, and their CF teams, understand whether the treatment is working and, if so, whether other treatments are still necessary.

With the benefit of machine learning, CF could be a less disruptive force in people's lives; they could know whether their treatments are effective or not and they could avoid cross-infection risks by cutting down unnecessary time in the clinic. More generally, people with CF will have the information they need to manage, and make better informed decisions about, their healthcare, their lives and their futures. The burden of treatment could lessen, both as a result of removing unnecessary treatment components and by the willpower-conserving impact of knowing whether specific treatments are really helping or not. As we all know, it is easier to do something unpleasant or inconvenient if you know it is going to help you.

A model of care like this could also help demonstrate the economic case for the high-priced disease-modifying therapies that typically target specific genotypes within the population of people with CF. It could lighten the load on the CF healthcare system as the population of people with CF grows due to better survival.

14 References

1 Eric Topol. The Patient Will See You Now. Basic Books, 2015.

2 https://www.macmillan.org.uk/_images/cancer-statistics-factsheet_tcm9-260514.pdf

3 https://www.who.int/news-room/fact-sheets/detail/cancer

4 Ahmed Alaa, Mihaela van der Schaar, "Prognostication and Risk Factors for Cystic Fibrosis via Automated Machine Learning," *Scientific Reports,* 2018.

5 Bryan Lim, Mihaela van der Schaar, "Disease-Atlas: Navigating Disease Trajectories using Deep Learning," *Machine Learning for Healthcare Conference (MLHC),* 2018.

6 Ahmed Alaa and Mihaela van der Schaar, "AutoPrognosis: Automated Clinical Prognostic Modeling via Bayesian Optimization with Structured Kernel Learning," ICML, 2018.

7 Jinsung Yoon, William Zame, Abijit Banerjee, Martin Cadeiras, Ahmed Alaa and Mihaela van der Schaar, "Personalized survival predictions via Trees of Predictors: An application to cardiac transplantation," PloS One, 2018.

8 Jinsung Yoon, William Zame and Mihaela van der Schaar, "ToPs: Ensemble Learning with Trees of Predictors," IEEE Transactions on Signal Processing (TSP), 2018.

9 Owen Lahav, Nicholas Mastronarde, and Mihaela van der Schaar, "What is Interpretable? Using Machine Learning to Design Interpretable Decision-Support Systems," *NIPS Machine Learning for Health Workshop* 2018.

10 M. Ribeiro, S. Singh and C. Guestrin. Why should I trust you? Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD

11 Jinsung Yoon, James Jordon, Mihaela van der Schaar, "RadialGAN: Leveraging multiple datasets to improve target-specific predictive models using Generative Adversarial Networks," *ICML, 2018.*

12 Ahmed Alaa, Mihaela van der Schaar, "Deep Multi-task Gaussian Processes for Survival Analysis with Competing Risks," *NIPS, 2017*

13 Alexis Bellot, Mihaela van der Schaar, "Multitask Boosting for Survival Analysis with Competing Risks," *NIPS, 2018*

14 Changhee Lee, William Zame, Jinsung Yoon and Mihaela van der Schaar, "DeepHit: A Deep Learning Approach to Survival Analysis with Competing Risks," AAAI, 2018.

15 Geert Litjens, et al. "A survey on deep learning in medical image analysis." Medical image analysis 42: 60-88, 2017.

16 Varun Gulshan, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA, 316(22):2402–2410, 2016.

17 Jie-Zhi Cheng, et al. Computer aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans. Scientific Reports, 6:24454, 2016.

18 Mark Cicero, et al. Training and validating a deep convolutional neural network for computer-aided detection and classification of abnormalities on frontal chest radiographs. Investigative Radiology, 52(5):281–287, 2017.

19 Thijs Kooi, et al. Large scale deep learning for computer aided detection of mammographic lesions. Medical Image Analysis, 35:303–312, 2017.

20 Andre Esteva, et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639):115–118, 2017

21 H A Haenssle, et al. Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of Oncology, page mdy166, 2018.

22 Ahmed Alaa, Kenneth Moon, William Hsu and Mihaela van der Schaar, "ConfidentCare: A Clinical Decision Support System for Personalized Breast Cancer Screening," *IEEE Transactions on Multimedia – Special Issue on Multimedia-based Healthcare,* 2016

23 Stefan Wager and Susan Athey. Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests. arXiv:1510.04342, 2017.

24 Ahmed Alaa and Mihaela van der Schaar, "Limits of Estimating Heterogeneous Treatment Effects: Guidelines for Practical Algorithm Design," ICML, 2018

25 Ahmed Alaa and Mihaela van der Schaar, "Bayesian Nonparametric Causal Inference: Information Rates and Learning Algorithms," accepted and to appear in IEEE Journal of Selected Topics in Signal Processing (JSTSP), 2018.

26 K. Henry, D. Hager, P. Pronovost and S. Saria. A targeted real-time early warning score (TREWScore) for septic shock. Science Translation Medicine. 7(299)299ra 122, 2015

27 Jinsung Yoon, Ahmed Alaa, Scott Hu, Mihaela van der Schaar, "ForecastICU: A Prognostic Decision Support System for Timely Prediction of Intensive Care Unit Admission," ICML 2016.

28 Ahmed Alaa, Jinsung Yoon, Scott Hu, and Mihaela van der Schaar, "Individualized Risk Prognosis for Critical Care Patients: A Multi-task Gaussian Process Model," Big Data in Medicine: Tools, Transformation and Translation, Cambridge. 2017.

29 Ahmed Alaa, Jinsung Yoon, Scott Hu and Mihaela van der Schaar, "Personalized Risk Scoring for Critical Care Patients using Mixtures of Gaussian Process Experts," ICML 2016 – Workshop on Computational Frameworks for Personalization.

30 Brian G Arndt, et al. Tethered to the EHR: Primary care physician workload assessment using EHR event log data and time-motion observations. The Annals of Family Medicine, 15(5):419–426, 2017.

31 Christopher D Manning and Hinrich Schütze. Foundations of statistical natural language processing. MIT press, 1999.

32 Cem Tekin, Onur Atan and Mihaela van der Schaar. Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis. *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 2, pp. 220 – 234, 2015.

33 Luca Canzian, Yu Zhang, Mihaela van der Schaar, "Ensemble of Distributed Learners for Online Classification of Dynamic Data Streams," *IEEE Trans. on Signal and Information Processing over Networks, vol. 1, no. 3,* 2015

34 Yannick Meier, Jie Xu, Onur Atan, and Mihaela van der Schaar, "Predicting Grades," *IEEE Transactions on Signal Processing*, Feb. 2016.

35 Cem Tekin, Onur Atan and Mihaela van der Schaar, "Discover the Expert: Context-Adaptive Expert Selection for Medical Diagnosis," *IEEE Transactions on Emerging Topics in Computing*, June 2015

36 David Silver, et al. Mastering the game of go with deep neural networks and tree search. Nature, 529(7587):484–489, 2016.