



Contents lists available at ScienceDirect

Journal of Cystic Fibrosis

journal homepage: www.elsevier.com/locate/jcf

Opportunities for machine learning to transform care for people with cystic fibrosis



1. Towards an Individualized Approach to CF Patient Care

The availability of high-quality data from patient registries provides a robust starting point for using Machine Learning (ML) techniques to enhance the care of the patient with cystic fibrosis (CF). Capitalizing on the wealth of information provided by registry data, ML techniques can augment clinical workflows by making individual-level predictions for a patient's prognosis that are tailored to their specific traits, features, and medical history. Such personalized approaches become especially relevant as CFTR modulators precipitate a shift to mutation-based medicine. ML-based techniques can help provide clinicians with a refined understanding of patient heterogeneity. Here, we discuss several areas where ML techniques can help underpin a personalized approach to patient management.

CF care often entails making a variety of (potentially life-changing) decisions at different stages of disease. These decisions should ideally be informed, among other things, by the individual patient's clinical history, the observed lifetime health trajectories of previous patients, and the similarities between the patient at hand and previous patients. Because CF is a complex disease with a myriad of possibilities and uncertainties about the prognoses of individual patients, domain knowledge about the "typical" CF patient and anecdotal experiences of individual clinicians, while critical, cannot account for the vast phenotypic diversity of the patient population. From a patient's perspective, CF management entails a gruelling daily burden of self-administered treatment. While there have been major advances in CF therapeutics, each new treatment has typically been added to the existing regimen. The polypharmacy aspect of CF makes it hard to decide whether a particular treatment is necessary for an individual at a given time, i.e. whether new treatments are helping and whether existing treatments are still bringing therapeutic value. This model of care can become a burden in itself, and many of these hurdles might be mitigated by moving from a one-size-fits-all model of care to a more individualized one.

ML technology has demonstrated a great ability to extract actionable intelligence from data in a variety of application domains [1]. ML models can address patient heterogeneity by learning the underlying complex patterns that govern how individual patient features and traits map to different prognoses. ML-based models can guide personalised decisions through individual-level predictions on: whether the patient at hand is likely to have an exacerbation, whether a specific new treatment is likely to be effective for that patient, what is the most likely sequence of clinical events that the patient might experience, and what is the rela-

tive likelihood of their different competing risks (Fig. 1). Moreover, ML-based models trained on patient-generated data recorded on digital devices, in the context of macro-intelligence extracted from registries, can inform whether a proactive clinical intervention is needed, and even whether a clinic visit is necessary. In a world like this, CF could be a less disruptive force in people's lives, individuals with CF would know that they are using only those treatments that are actually helping them. More generally, they will have the information they need to manage and make better informed decisions about their healthcare, their lives and their futures. A model of care like this could help inform the health economics of the high-priced disease-modifying therapies that target specific genotypes within the population of people with CF. It could also lighten the load on the CF healthcare system as the population of patients grows due to better survival.

2. What can Machine Learning Offer?

In what follows, we discuss various areas where ML can be deployed to enhance patient care and introduce some of the work that has already produced encouraging results. This opportunity is now available in part because of the CF data registries. These registries that are now available in several countries (e.g. UK, US, France, Canada) provide an abundance of data including many variables and longitudinal biomarkers for several thousands of patients over long periods of time. ML can leverage these data to enable personalised patient care in the following ways:

2.1. Risk prediction

A CF patient faces a variety of health risks. For instance, infections or deterioration of lung function. Currently, risk evaluation for CF patients, and more generally in the medicine domain, does not fully integrate the wealth of information available about the patient. It is usually done on the basis of a linear model, such as a Cox proportional hazards model [2] (e.g. see [3]), and using a relatively small number of hand-picked features, often chosen by clinical judgment. For example, in this issue, Juarez-Colunga et al. develop a Cox regression model to predict pulmonary exacerbations and infer the associated, population-level risk factors [4]. ML methods would allow us to take such work one step forward by learning patient-specific information that might influence treatment decision – recent methods has been proposed to extract patient-specific risk factors from ML models using symbolic regression approaches [5].

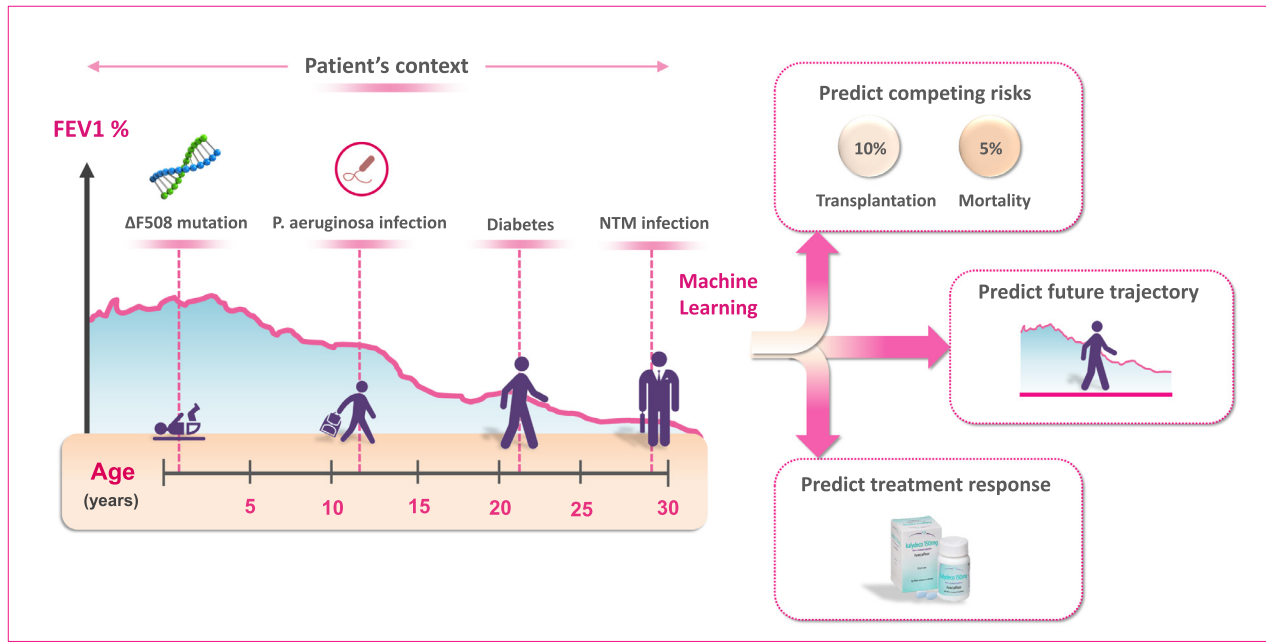


Fig. 1.

In certain situations, 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. 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 yield more accurate predictions, and hence better treatment guidance, for the patient at hand.

An example of risk evaluation in CF is the decision on lung transplant (LT) referral for patients with advanced disease. Currently, the recommendations are to refer patients based upon their forced expiratory volume (FEV_1). Patients with FEV_1 below 30% of its predicted value are recommended for referral for LT, even earlier with other risk factors [6]. AutoPrognosis [7] is a recently proposed ML method that has been proposed to predict patients' need for LT. While AutoPrognosis found that FEV_1 is indeed the single most important clinical variable, it also provided crucial insights into the importance of other variables in making accurate predictions. For example, oxygenation: variables that reflected disorders in gas exchange in the lungs played a key role in improving the precision, and therefore the usefulness, of the prognostic models. Using data from the UK CF registry, the ML method has demonstrated a 35% improvement in accuracy over traditional methods for clinical referral [7].

Another aspect of AutoPrognosis is that it provides an automated process for applying ML to different problems. 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. AutoPrognosis uses ML itself to both choose and tune the ML-model which works best for the particular task and data at hand thereby making ML more accessible to clinical researchers with little or no technical knowledge of ML. Moreover, using AutoPrognosis, we can build an ML model that is continually evolving as more data are collected over time.

2.2. Longitudinal trajectories

In CF, biomarkers and other risk factors are measured repeatedly over time. ML can process the longitudinal trajectory of

biomarkers and help clinical decision-makers better understand the disease and predict its trajectory in individuals. There has long been interest in the value that imaging brings to patient care and Brasfield chest radiographic scoring has been used in research but never has found a home in the clinic. In this issue Zucker et al. [8] used a deep neural network model to automate Brasfield scoring and reduce the burden on pediatric radiologists, but much work needs to be done to demonstrate real-time clinical utility in disease monitoring. Similarly, a new model used a variation of neural networks to capture the complex dependencies among different features of the patient over time [9]. The model was tested on the UK CF registry, and was able to accurately predict CF progression trajectories and also provide insights into the dynamics of disease progression.

2.3. Competing risks

In CF patients suffer from, or are at risk of, 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. Evaluating these probabilities of competing events in a personalized manner is a challenging task. Recently, a dynamic prediction model was proposed [10] which uses deep learning for evaluating the probability of different competing risks using longitudinal biomarkers and produces accurate predictions of how competing risks evolve over time. Deep learning has been successful in recent years among different ML algorithms and part of the success of the method in [10] is due to the capability of deep learning to capture complex interaction of underlying disease progression.

2.4. Personalized treatment recommendations

Among the available treatment options, ML can help predict which one is more suitable for an individual with CF given their particular characteristics. This can improve outcomes for patients who are prescribed the optimal treatment regimen and improve utilization of healthcare resources through avoidance of treatments that will yield little or no additional benefit for individuals.

There are multiple ML works in this area, a prominent example is [11].

2.5. Referrals to hospital, remote monitoring, and early warnings

As we discussed in the previous section, CF patients need to routinely make clinical visits even when well, which is not an efficient model. ML, with the use of digital devices, can transform this model of care. Digital sensors and wearable devices can enhance care in (at least) three ways: 1- By collecting information that can be used by the treating clinicians (especially in tracking patients' key health indicators and progress). 2- By collecting information that can be used by the patient to monitor and manage their own progress. 3- By providing actionable intelligence to the patients including early warnings, reminders, and suggestions for achieving success with medication, exercise and treatment regimens [12]. Remote monitoring enables a feedback loop that may enrich the quality of the clinician-patient relationship even though the patient may make fewer visits to the clinic. ML will play a vital role in this by creating systems that can: integrate all the relevant data; make predictions about the patients' health, provide feedback about the patients' progress; offer recommendations including alternative regimes of diet and exercise; and alert the patient and medical personnel when further action or consultation is needed.

2.6. Scientific discovery

An additional advantage of ML-based models is that since they are capable of integrating many features, and capturing complex patterns, they make it possible to discover the clinical significance of specific features that were not previously understood to be important. For example, [7] reveals an unexpectedly important role for supplemental oxygenation, in addition to FEV₁, in predicting health decline for CF patients.

3. Challenges and the Way Forward

We believe that ML can stimulate a patient-centred revolution in healthcare, particularly in complicated diseases like CF, by enabling and empowering clinicians and researchers to extract more value from existing and emerging health data streams. It can 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. Critically, ML will support and complement, rather than substitute for, the judgment of medical professionals, it will empower patients and inform health administrators. The purpose of ML in the medical domain is to provide actionable intelligence and decision support to all the stakeholders. In order to realise the full potential of ML to enhance CF care, it is vital that CF clinicians and other specialist healthcare professionals understand the capabilities and limitations of ML and play a pro-active role in the development and application of ML to create clinically valuable tools in the unique domain of CF care. At the same time, it is critical that researchers and clinicians work with people with CF to develop tools that lighten their burden; make it easier for them to manage healthcare decisions; and achieve the best possible outcomes.

Acknowledgments

The authors would like to thank Dr. Janet Allen (Director of Strategic Innovation, UK Cystic Fibrosis Trust) and the UK Cystic Fibrosis Trust for supporting their research.

Mahed Abroshan

*The Department of Applied Mathematics and Theoretical Physics and
The Department of Public Health, University of Cambridge,
United Kingdom*

Ahmed M. Alaa

*Department of Electrical and Computer Engineering, University of
California Los Angeles, Los Angeles, California, USA*

Oli Rayner

Person with CF, Plymouth, United Kingdom

Mihaela van der Schaar*

*The Department of Applied Mathematics and Theoretical Physics and
The Department of Public Health, University of Cambridge,
United Kingdom*

*Department of Electrical and Computer Engineering, University of
California Los Angeles, Los Angeles, California, USA*

*Corresponding author.

E-mail address: mihaelacla@gmail.com (M. van der Schaar)

References

- [1] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015;521(7553):436.
- [2] David CR. Regression models and life tables (with discussion). *J Roy Stat Soc* 1972;34(2):187–220.
- [3] Keogh RH, Seaman SR, Barrett JK, Taylor-Robinson D, Szczesniak R. Dynamic prediction of survival in cystic fibrosis: A landmarking analysis using UK patient registry data. *Epidemiology (Cambridge, Mass.)*. 2019; Jan; 30(1):29–37. doi:10.1097/EDE.0000000000000920.
- [4] Juarez-Colunga E, Rosenfeld M, Zemanick ET, Wagner B. Application of multiple event analysis as an alternative approach to studying pulmonary exacerbations as an outcome measure. *J Cyst Fibr* 2020;19:114–18.
- [5] Alaa AM, van der Schaar M. Demystifying Black-box Models with Symbolic Metamodels. In: *Advances in Neural Information Processing Systems*; 2019. p. 11301–11.
- [6] Ramos KJ, Smith PJ, McKone EF, Pilewski JM, Lucy A, Hempstead SE, Tallarico E, Faro A, Rosenbluth DB, Gray AL, Dunitz JMCF Lung Transplant Referral Guidelines Committee. Lung transplant referral for individuals with cystic fibrosis: Cystic Fibrosis Foundation consensus guidelines. *J Cyst Fibr* 2019;18(3):321–33.
- [7] Alaa AM, van der Schaar M. Prognostication and risk factors for cystic fibrosis via automated machine learning. *Sci Rep-UK* 2018;8(1):11242.
- [8] Zucker EJ, Barnes ZA, Lungren MP, Shpanskaya Y, Seekins JM, Halabi SS, Larson DB. Deep learning to automate Brasfield chest radiographic scoring for cystic fibrosis. *J Cyst Fibr* 2020;19:131–8.
- [9] Alaa AM, van der Schaar M. Attentive state-space modelling of disease progression. In: *Advances in Neural Information Processing Systems*; 2019. p. 11334–44.
- [10] Lee C, Yoon J, Van Der Schaar M. Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis with Competing Risks based on Longitudinal Data. *IEEE T Bio-Med Eng* 2020; Jan; 67(1):122–33. doi:10.1109/TBME.2019.2909027.
- [11] Alaa AM, van der Schaar M. Bayesian inference of individualized treatment effects using multi-task gaussian processes. In: *Advances in Neural Information Processing Systems*; 2017. p. 3424–32.
- [12] Davies SC. Annual Report of the Chief Medical Officer, 2018, Health 2040 Better Health Within Reach. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/767549/Annual_report_of_the_Chief_Medical_Officer_2018_-_health_2040_-_better_health_within_reach.pdf