Health Informatics via Machine Learning for the Clinical Management of Patients

D. A. Clifton, K. E. Niehaus, P. Charlton, G. W. Colopy
Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford, UK

Abstract
The modern generation of health informatics systems based on machine learning incorporates a range of technologies from (i) wearable sensors for the acquisition of data from patients, through to (ii) biomedical signal processing methods for conditioning the high-resolution data that result, through to (iii) “big data” machine learning approaches to fuse the heterogeneous data that are currently collected from routine clinical care in many healthcare systems. The latter may include fusion of time-series data from sensors; categorical data from electronic health records (EHRs); and, increasingly, biomarkers derived from genomics, proteomics, and other sources. While this field holds substantial promise for the future of medicine, and for our ability to tailor care to the particular physiology of the individual, the penetration of such systems based on machine learning into actual clinical practice is in its infancy. This review surveys the recent literature in this rapidly-changing field, aiming to investigate how health informatics systems that employ machine-learning methods are affecting the clinical management of patients. While the range of literature presented is broad, underpinning each publication is the demonstration of machine learning methods for health informatics.

This review describes health informatics systems that are based on machine learning, throughout the patient journey through a typical hospital healthcare system, from (i) the intensive care unit (ICU), to (ii) discharge and subsequent monitoring on acute wards and on general wards, to (iii) wider-scale tracking of patient condition using the EHR.

1 Clinical Management in the Intensive Care Unit
The ICU routinely cares for the most severely ill patients in the hospital, allowing for the provision of radical lifesaving treatment. Once admitted, the condition of patients is heavily monitored to facilitate the restoration of their physiology to “normality”. Many clinically-relevant parameters, such as heart rate (HR) and blood pressure (BP), are monitored as frequently as every second. However, due to the limitation of human resource allocation, much of this information is not used by clinical staff. Furthermore, critical care medicine is one of the fastest-growing medical fields in terms of patient numbers [1], and the complexity of critical illness requires interpretation of numerous interacting parameters. As a result, the ICU has become a promising area for the development of novel informatics for the real-time processing of streams of physiological data, with the aim of providing clinical decision support for patient management.

Prediction of patient outcomes is an immediately obvious application of health informatics and has a long history, which has formed a basis for more recent innovations. In 1981, Knaus et al. published the acute physiology score (APS) as a component of the acute physiology, age, and chronic health evaluation (APACHE) system [2]. With the recent increase in the quantity of available data, and the advances in machine learning techniques that allow these data to be processed, there is a great opportunity for creating accurate patient-specific prediction models, which therefore allow an assessment of individual patients, suitable for affecting management decisions. Recent work
by Johnson et al. applied a hybrid system combining a genetic algorithm with particle swarm optimisation to create an ICU severity score that used a parsimoniously-selected feature set [3]. However, these models use aggregate summaries of patient physiology over the first hour or day in the ICU (such as the maximum HR). Constraining the analysis to considering only early-admission data in this way is, as the authors describe, due to the traditional setting of ICU severity scores for use at admission.

A recent review of all mortality prediction studies found 94 investigations that contained 118 predictive tools [4]. Of these 118 tools, 34 were unique. This demonstrated that there is a lack of wide uptake of more sophisticated mortality prediction models. However, there exist promising approaches undertaken in recent years. Saria et al. present a model which aimed to predict neonatal morbidity using continuous vital-sign monitoring data from the first 3 hours of the neonate’s life [5]. The model first filtered the continuous vital-sign data into long- and short-term variability signals. For each class of patient, a probability distribution for each vital sign was estimated, and the probabilities of morbidity from each of these distributions were aggregated into a single score using regularised logistic regression. The resulting score, named the PhysiScore, had excellent discrimination for various neonatal morbidities as compared to more commonly utilized severity scores. To ascertain whether the dynamics of the heart rate could further inform a predictive model, the same authors then apply an extension of Latent Dirichlet Allocation (LDA) for time-series data. The model creates “topics” for each time series which are shared across patients. These topics refer to concepts such as “healthy” or “lung”, and control the distribution of autoregressive models which describe the data. The result was validated with 30 infants and demonstrated that the length of time a neonate’s condition was deemed to spend in the “healthy” topic was correlated with severity of that neonate’s condition.

Similar models have been used to monitor adult ICU patients. Lehman et al. apply a switching vector autoregressive (SVAR) model to the dynamics of HR and BP time-series for patients in the ICU [6]. This model similarly summarises the time-series using autoregressive modes and uses the proportion of time in each mode as a marker of health or severity. The authors showed that BP dynamics alone are able to outperform existing ICU scores. Furthermore, they show that the addition of BP dynamics to the widely-used APACHE IV score further improved its AUROC from 0.82 to 0.85.

One particularly successful health informatics system is the HeRO monitor, which has been validated via a randomised controlled trial [7]. This system provides early prediction of neonatal sepsis by incorporating various measures of HR variability including the standard deviation of R-R Intervals, sample asymmetry, and sample entropy. These metrics captured a common pattern in HR observed in neonates with sepsis: rapid decelerations combined with low variability. A comparison of the control and study population demonstrated that for every 48 low birth weight infants, the HeRO monitor saved one life.

In addition to the numerous time-series signals acquired in the ICU, there also exists a large body of useful information contained within the clinical notes written during a patient’s stay in the ICU. These notes usually contain information regarding past medical history, diagnostic information, current patient status, treatment plans and discussion of end-of-life care options. These notes represent an underused, but highly relevant, source of information, and are now often recorded electronically, making them more suitable for automated interpretation. Ghassemi et al. demonstrated the utility of incorporating information contained in these notes into ICU-based systems [8]. The system extracts a set of 50 topics from the case notes for patients, using an LDA model derived from the top 500 words from the notes. The ability of the model to predict long-term mortality increased AUROC from 0.711 to 0.788 when such data were included in the ICU’s severity score. Further research by the same authors demonstrated that similar performance could be achieved by an algorithm implemented at the bedside [9]. The AUROC of the latter model for predicting mortality (without information from the ICU notes) was 0.76, while the addition of learnt topics from the notes improved this to 0.84.

2 Clinical Management Outside the ICU: Wearable Sensors

The majority of patients in hospitals are typically ambulatory, and in this setting patient condition is often monitored only by manual observation by the clinical staff. It is established that the use of wearable sensors for managing the care of such populations can improve patient outcomes. A recent study by Cleverley et al. [10] have found that remote monitoring is associated with improved survival to hospital discharge, following in-hospital cardiac arrest. However, despite such studies, evidence suggests that the current use of systems based on wearable sensors in hospital is far from ideal [11]. This has been attributed to systems being used inappropriately, in an attempt to manage patients for whom such monitoring confers no benefit [12,13]. This latter effect has been attributed to the inadequacy of current guidelines for stratifying patients to determine whether or not additional monitoring will be of benefit [14].

The standard of clinical management for patients on acute hospital wards is typically to have vital signs measured manually every 4-6 hours [15]. There is evidence that wearable sensors are being used in clinical practice to increase the frequency of measurement of the vital signs, with the aim of identifying the deteriorating patient [16-18]. Automated analysis of the trends of vital signs acquired from wearable sensors has been used in the care of ambulatory post-operative patients with the aim of determining if the time-series data from a patient appear similar to those of “normal” or “at risk” patients [19-21]. This latter study used Bayesian Gaussian processes (GPs) to perform clustering of entire univariate time-series; a similar approach was described by Duerichen et al. [22], in which multiple time-series of vital signs were modelled simultaneously using multi-task Gaussian processes, for the monitoring of ambulatory patients. Vital-sign data from wearable sensors were fused with the manual observations made by the clinical staff for acutely ill, ambulatory patients in Clifton et al. [23-
25], using GPs and one-class support vector machines. Such methods have been extended by the use of health informatics based on extreme value theory, a branch of statistics used to model extremal data, and which have been demonstrated in a patient monitoring study in a US hospital [26, 27]. The evidence base for the use of health informatics systems based on machine learning was questioned in [28], where methods for evaluating the efficacy of a system used for identifying the deteriorating patient in a large Emergency Department setting were described.

Gaussian processes are non-parametric Bayesian methods of performing time-series analysis, which are becoming increasingly adopted by probabilistic health informatics systems based on machine learning. They offer a principled manner of coping with noisy or incomplete data, as typically encountered in routine clinical management. We note in passing that further applications of this approach outside the realm of sensor-based analysis include the assessment of heat-exposure sickness [29, 30], malaria infections [31], and the management of patient billing records [32].

Other recent studies, such as [33], focussed on using existing sensors (such as ECG electrodes) for the assessment of respiratory rate (RR) from cohorts of similarly ambulatory patients. This latter area of application has also been addressed by the use of pulse oximeters, again worn by ambulatory patients [34-36]. The penetration of such systems into actual clinical management is limited, and this is a point addressed by Orphanidou et al. [37] in which algorithms were developed to improve both the robustness of RR estimation from pulse oximeters, and to extend their battery life such that they could be used in the clinic without frequent replacement of batteries. The use of pulse oximeters to estimate other vital signs, such as changes in BP from the pulse transit time (the interval between a QRS complex from the ECG and the corresponding pulse of a finger-worn oximeter), has been well-described in the literature, but a recent study [38] claims to have improved the perceived lack of robustness that has previously prevented the use of such methods in clinical practice.

3 Clinical Management Outside the ICU, via “Big Data” Health Informatics

Beyond the traditional monitoring of vital signs, there is a rich resource of data collected from the routine care of patients in many hospitals. This “big data” problem typically means that only specific subsets of data are used for clinical management, whereas health informatics systems based on machine learning offer the potential for informing clinical care by fusing information from the EHR. This section considers those EHR-based analyses that pertain to applications outside the ICU (where the latter were described in section 1).

Two prominent themes within health informatics concern the integration of genomic data into patient care and the intelligent use (and re-use) of EHR data [39-41]. While research output in these areas is increasing, translation into routine clinical practice is still slow. As Bright et al. report in their review of clinical decision-support systems, while there are many small-scale examples of local adoption of clinical support tools, there is still a dearth of evidence on multi-site clinical outcomes [42]. There are also few examples of predictive systems, even in the research stage, that employ both clinical and genomic features. Partially, this is due to our still rudimentary understanding of human genomics: in 2014, Dewey et al. reported on their experience of whole-genome sequencing (WGS) twelve healthy subjects [43]. The predominant finding was the WGS is not yet ready for use in clinical practice, given the inconsistency of results, inadequate base overage in key disease-related areas of the genome, and rudimentary understanding of how the majority of SNPs may affect the risk of certain diseases. Much informatics research, therefore, is focused upon developing tools that will enable us to answer more fundamental questions of human physiology. EHR phenotyping algorithms, which are designed to enable large-scale genome-wide association studies (GWAS), fall into this category. Though this research is not intended to produce immediate clinical decision support tools, it will lead to the development of appropriate therapies and diagnostic tools further down the line.

The EHR is a rich source of data for many avenues of research outside the ICU, providing input data for both traditional epidemiology studies and training data for the development of predictive algorithms. However, patient records are often partially missing, incorrect, systematically misleading, and contradictory. Particularly relevant for supervised machine learning analysis, patient labels may be confounded by the fact that care is being provided [44]. Hripcsak and Albers [45] outline several of the associated challenges of extracting EHR data that reflects true patient physiology.

The Electronic Medical Records and Genomics Network (eMERGE) is a large project funded by the National Human Genome Research Institute in the US, with the goal of hastening the use of genome sequencing in clinical practice. The project’s first phase (eMERGE-I) formed a consortium of five large clinical sites, each with access to both EHR and biobank data. The primary goal was the establishment of EHR-phenotyping algorithms, to enable GWAS by linking large EHR-phenotyped patient cohorts with their genomic biobank data [46]. The eMERGE-I project was quite successful in extracting phenotypes from the EHR, publishing algorithms for phenotypes ranging from rheumatoid arthritis to drug-induced liver disease. These algorithms are publicly-available and maintained at the Phenotype KnowledgeBase (PhKB). In the second phase of eMERGE funding (eMERGE-II), the focus has shifted towards implementing genomic tools as clinical support systems in EHRs. The PGx project is the first of these translational informatics projects within the eMERGE network, and aims to implement a system that will use knowledge of pharmacogenomic relationships to alert physicians when a patient may require an adjusted medication or dosage [47].

This research is not confined to the eMERGE network; for example, Wang, et al. used semi-supervised machine learning to extract ovarian cancer diagnoses from the EHR [48]. Shivade et al. recently conducted a review of phenotyping algorithms, finding 97 studies published between 2010-2013. Many of these have used natural-language processing (NLP) and rule-based defini-
tions in their phenotype definitions, often followed by machine learning or statistical analyses on the resulting phenotype-extracted patients’ data [49].

The integrating biology and the bedside (i2b2) initiative [50], the UK Biobank, and the Kaiser Permanent Research Program on Genes, Environment, and Health (RPGEH) also all similarly aim to combine large amounts of clinical and genomic data; Jensen et al. provide additional examples [40]. The EHR-based phenotyping efforts described above have largely been focused upon identifying patients who can be described by current disease terminology. However, with growing access to EHR data, it is now becoming possible to identify novel patient subgroups. While many complex diseases have traditionally been thought of as monolithic entities, it is very likely that finer-grained phenotypic sub-types exist [51,52]. Ho et al. have framed the problem in terms of tensors, using nonnegative tensor factorisation to more automatically identify patient phenotypes [53].

Doshi-Velez et al., for example, looked at the timing and presence of ICD9 diagnostic codes in the EHRs of over 13,000 autism spectrum disorder patients [54]. Using hierarchical clustering, they found three distinct patient trajectories, corresponding to a group of patients characterised by seizures, a group by multisystem disorders, and a group by psychiatric disorders. Schulam, et al. built upon such approaches by developing a probabilistic model to cluster patient trajectories of various clinical indicators related to scleroderma [55]. Lasko et al. used Gaussian process regression over serum uric acid measurements, combined with a two-layered neural network, to create a sparse feature set of time-series trajectories. When projected into two dimensions, the learned feature sets showed good discrimination between known diseases and also suggested possible additional underlying disease subtypes [56]. The goal of such research is primarily hypothesis-generating, with the aim of determining if patients within phenotypic subgroups are united by a common genomic pathway, or if they might respond to certain types of treatment. Such analysis is not intended to be immediately translatable into the clinic, but informs our clinical understanding of these disorders and points towards promising future areas of research.

While the studies presented above enable the study of physiological mechanisms that may be involved in human disease, there are many examples of recently-developed systems that are designed to be incorporated into an intelligent EHR system. Huang et al. used features derived from the EHR to create a screening tool for depression with sensitivities and specificities on par with those achieved by primary care physicians [57]. Wiens et al. created a risk-stratification model for becoming infected with Clostridium difficile by incorporating the time-based relationship between clinical variables in the EHR; the resulting system was able to predict patient risk much more successfully than when considering features in aggregate [58]. Van der Heijden et al. used temporal Bayesian networks to predict exacerbation events for chronic obstructive pulmonary disease patients based on EHR data; their system was able to predict an exacerbation within 24 hours with an AUROC of 0.90 in a validation study [59].

On a larger scale, Li et al. illustrate how EHR data can be mined and combined with curated genomic data to produce early prognostic risk indicators for common medical conditions [60]. The goal of such studies is for the resulting tools to be adopted into clinical practice as new screening techniques. One example of the successful translation of a genomics-based informatics tool has been the development of gene-expression based algorithms for breast cancer prognosis. Van De Vijver et al. [61] used supervised and unsupervised machine learning techniques on gene expression data from breast cancer cells to provide large improvement over existing clinical prognostic markers. These techniques have now been developed into a commercial tool (MammaPrint, Agenda, Netherlands) that is used for breast cancer prognosis in the clinic. The visualisation of such complex data was addressed by Ding et al., who developed a visualisation platform using unsupervised learning methods [62]. Das et al. provide a review of such informatics techniques and data types that have been used for predicting cancer prognoses [63], with the aim of improving the clinical management and treatment of such patients.

Clinical microbiology is another field that is set to be transformed through genome sequencing and supportive informatics tools embedded within the EHR [64]. While genome sequencing, combined with machine learning algorithms, has been used for HIV drug resistance prediction for many years, similar methodologies are now being applied for bacterial infections. For instance, Stoesser et al. [65] and Gordon et al. [66] have illustrated the feasibility of using whole-genome sequences from Staphylococcus aureus, Klebsiella pneumonia, and Escherichia coli to determine bacterial drug resistance. Niehaus et al. illustrated the first steps towards creation of a machine-learning-based antibiotic resistance-prediction tool with the aim of providing appropriate treatments [67].

4 Conclusions

We conclude by emphasising that the field of health informatics systems based on machine learning, drawing on disparate datatypes from the ICU, the wider hospital, and from (potentially very complex) EHR data, is in its infancy. While the majority of hospitals in the developed world have implemented EHR systems of some kind, the integrated use of the large quantities of data that arise from such systems is not employed at scale. This article has surveyed recent developments in this field, in which the clinical management of patients has been affected by health informatics systems based on machine learning, or in which systems for performing such management are in development. We observe in closing that the barriers-to-entry for such activity are substantial, and are affected by the availability of multidisciplinary teams drawing on both clinical expertise and practitioners from the information sciences. The implications for research in this field are that clinical management of patients will become more dependent on machine-learning systems with the ever-increasing quantities of data that are being collected in routine care, from the ICU through to discharge, in most healthcare systems.
References

34. Wang Z, Shah AD, Tate AR, Denaxas S, Shawe-Taylor J, Hemingway H. Extracting diag...


