Available online at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/burns

Mortality risk prediction in burn injury: Comparison of logistic regression with machine learning approaches

Neophytos Stylianou [a,*], *Artur Akbarov* [a], *Evangelos Kontopantelis* [a], *Iain Buchan* [a], *Ken W. Dunn* [b]

[a] *Centre for Health Informatics, Institute of Population Health, University of Manchester, UK*
[b] *University Hospital South Manchester, Greater Manchester, UK*

ARTICLE INFO

Article history:
Accepted 28 March 2015

Keywords:
Machine learning
Burn
Mortality
Clinical prediction

ABSTRACT

Introduction: Predicting mortality from burn injury has traditionally employed logistic regression models. Alternative machine learning methods have been introduced in some areas of clinical prediction as the necessary software and computational facilities have become accessible. Here we compare logistic regression and machine learning predictions of mortality from burn.

Methods: An established logistic mortality model was compared to machine learning methods (artificial neural network, support vector machine, random forests and naïve Bayes) using a population-based (England & Wales) case-cohort registry. Predictive evaluation used: area under the receiver operating characteristic curve; sensitivity; specificity; positive predictive value and Youden's index.

Results: All methods had comparable discriminatory abilities, similar sensitivities, specificities and positive predictive values. Although some machine learning methods performed marginally better than logistic regression the differences were seldom statistically significant and clinically insubstantial. Random forests were marginally better for high positive predictive value and reasonable sensitivity. Neural networks yielded slightly better prediction overall. Logistic regression gives an optimal mix of performance and interpretability.

Discussion: The established logistic regression model of burn mortality performs well against more complex alternatives. Clinical prediction with a small set of strong, stable, independent predictors is unlikely to gain much from machine learning outside specialist research contexts.

1. Introduction

Death in hospital is the main outcome that is used to monitor the quality of specialist burn services. Such services systematically collect, collate and analyse data on patients they admit, usually contributing to national registries that monitor death rates and compare different centres. Comparisons need to be adjusted for risk factors such as the age of the patient and the total surface area of the body that has been burned.

* *Corresponding author at*: Centre for Health Informatics, Vaughan House, University of Manchester, Manchester M13 9GB, UK.
E-mail address: neophytos.stylianou@postgrad.manchester.ac.uk (N. Stylianou).

http://dx.doi.org/10.1016/j.burns.2015.03.016

In England and Wales the burn services contribute data to the international burn injury database (iBID), which has been collecting data since 2003, following the 2001 National Burn Care Review's recommendations [1]. A wide range of demographic and clinical variables are collected on each patient in a structured way, and the analyses are fed back to help improve services.

In-hospital mortality is the most important metric by which burn service performance is measured. Accurate and reliable mortality risk prediction should not only support this performance measurement but also illuminate unexplained mortality. In addition to using such models at the population-level to evaluate services, clinicians also use them to assess outcomes in front-line care, which sometimes affects the targeting of resources to those who are more likely to benefit. Both uses of mortality prediction models can be therefore seen as resource management.

Mortality prediction models in burn injury have existed since the mid-20th century [2]. Up until 1980 they contained as predictor variables just age and total burn surface area (TBSA). In 1981 the Abbreviated Burn Severity Score (ABSI) was published, which for the first time, considered inhalation injury as a substantial independent predictor of mortality [3]. Since then many other models have been put forward and a recent systematic review [2] showed more than 40 currently in use. The vast majority of these models used logistic regression for development and validation, and for ease of computation they also developed a scoring system. For example one of the most recent prediction models is the Belgian Outcome in Burn Injury (BOBI) – a logistic model with a scoring system, and one of the few to use nationwide data [4].

More data and computation intensive machine learning methods could be used to predict mortality, taking into account for example subtle non-linear interactions between predictors, which regression models 'average out' [5]. Machine learning uses algorithms to automatically extract model-like 'structure' information from a given set of data. In contrast to parametric methods, machine learning techniques do not assume a priori knowledge about the statistical distributions that govern the data. It is a branch of artificial intelligence which uses systems that 'learn' from data, but to many it is considered a "black box" approach that is opaque to clinical validation. Various machine learning techniques have been used extensively over the past decade to predict mortality, for example genetic algorithms, artificial neural networks, support vector machines and random forests [6].

In this study we compare a wide range of mortality prediction models, including logistic regression, artificial neural network, support vector machine, random forest and naïve Bayes using data for patients admitted to specialised burn services in England and Wales from 2003 to 2011.

Table 1 – General description of burn data used for the study.

Parameters	Count	Overall
Mortality		
Survived	65,764	98.73 (%)
Died	847	1.27 (%)
Age		
Mean		25.50
Standard deviation		23.50
Median		21.00
25th percentile		2.00
75th percentile		42.00
TBSA		
Mean		3.96
Median		1.50
25th percentile		0.50
75th percentile		4.00
Inhalation injury		
Present	776	1.16 (%)
Absent	65,835	98.84 (%)
Existing disorders		
<3	64,209	96.39 (%)
≥3	2402	3.61 (%)
Injury type		
Flame	11,430	17.20 (%)
Flash	5031	7.50 (%)
Contact	14,952	22.50 (%)
Scald	27,353	41.10 (%)
Chemical	4346	6.50 (%)
Other	3499	5.20 (%)

Table 2 – Mean and standard deviation of the performance metrics for the test sets of the different prediction methods and the different optimisation method.

Optimisation method	Performance metric	Mortality prediction methods				
		Random forest	Support vector machine	Artificial neural network	Logistic regression	Naive Bayes
Discrimination	AUC	0.945 (0.006)	0.967 (0.006)	0.974 (0.004)	0.971 (0.005)	0.970 (0.004)
Sensitivity = Specificity	Sensitivity	0.83 (0.027)	0.917 (0.016)	0.922 (0.013)	0.919 (0.017)	0.917 (0.017)
	Specificity	0.967 (0.003)	0.920 (0.005)	0.934 (0.004)	0.923 (0.005)	0.926 (0.005)
	PPV	0.245 (0.012)	0.129 (0.006)	0.153 (0.006)	0.133 (0.006)	0.139 (0.006)
	Youden's	0.797 (0.025)	0.837 (0.016)	0.856 (0.011)	0.842 (0.012)	0.843 (0.013)
Sensitivity = PPV	Sensitivity	0.55 (0.032)	0.574 (0.038)	0.598 (0.039)	0.579 (0.035)	0.495 (0.034)
	PPV	0.55 (0.043)	0.566 (0.032)	0.599 (0.038)	0.568 (0.029)	0.491 (0.023)
	Specificity	0.994 (0.001)	0.994 (0.001)	0.995 (0.001)	0.994 (0.001)	0.993 (0.000)
	Youden's	0.544 (0.032)	0.568 (0.038)	0.592 (0.04)	0.574 (0.035)	0.489 (0.034)
Max Youden's Index	Youden's	0.762 (0.033)	0.834 (0.016)	0.856 (0.008)	0.839 (0.013)	0.841 (0.013)
	Sensitivity	0.786 (0.035)	0.904 (0.023)	0.928 (0.014)	0.907 (0.014)	0.919 (0.024)
	Specificity	0.976 (0.004)	0.930 (0.01)	0.928 (0.007)	0.93 (0.011)	0.922 (0.013)
	PPV	0.298 (0.03)	0.145 (0.016)	0.144 (0.011)	0.150 (0.016)	0.134 (0.017)