Journal of Biomedical Informatics 58 (2015) 133–144

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Prognostic models based on patient snapshots and time windows: Predicting disease progression to assisted ventilation in Amyotrophic Lateral Sclerosis

André V. Carreiro [a,*], Pedro M.T. Amaral [a], Susana Pinto [b], Pedro Tomás [a], Mamede de Carvalho [b], Sara C. Madeira [a,*]

[a] INESC-ID Lisbon and Instituto Superior Técnico, Universidade de Lisboa, Portugal
[b] Translational Clinical Physiology Unit, Institute of Molecular Medicine, Faculty of Medicine, Universidade de Lisboa, Portugal

ARTICLE INFO

Article history:
Received 19 February 2015
Revised 18 September 2015
Accepted 23 September 2015
Available online 8 October 2015

Keywords:
Prognostic model
Disease progression
Amyotrophic Lateral Sclerosis
Time windows
Patient snapshots

ABSTRACT

Amyotrophic Lateral Sclerosis (ALS) is a devastating disease and the most common neurodegenerative disorder of young adults. ALS patients present a rapidly progressive motor weakness. This usually leads to death in a few years by respiratory failure. The correct prediction of respiratory insufficiency is thus key for patient management. In this context, we propose an innovative approach for prognostic prediction based on patient snapshots and time windows. We first cluster temporally-related tests to obtain snapshots of the patient's condition at a given time (patient snapshots). Then we use the snapshots to predict the probability of an ALS patient to require assisted ventilation after k days from the time of clinical evaluation (time window). This probability is based on the patient's current condition, evaluated using clinical features, including functional impairment assessments and a complete set of respiratory tests. The prognostic models include three temporal windows allowing to perform short, medium and long term prognosis regarding progression to assisted ventilation. Experimental results show an area under the receiver operating characteristics curve (AUC) in the test set of approximately 79% for time windows of 90, 180 and 365 days. Creating patient snapshots using hierarchical clustering with constraints outperforms the state of the art, and the proposed prognostic model becomes the first non population-based approach for prognostic prediction in ALS. The results are promising and should enhance the current clinical practice, largely supported by non-standardized tests and clinicians' experience.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease characterized by a rapidly progressive muscular weakness. It causes denervation of axial, bulbar and respiratory muscles. This leads to a progressive functional impairment (in general without major cognitive decline [1]), and ultimately death [2]. It has no cure and its causes are yet to be discovered. Maintaining the patients' quality of life is thus of major relevance.

Respiratory complications account for the majority of deaths in ALS. Most patients succumb from hypoventilation with hypoxemia and hypercapnia, often associated with respiratory infection [3]. Predicting the onset of hypoventilation is therefore of major importance, to anticipate timely interventions such as the start of non-invasive ventilation (NIV). NIV was shown to be effective in improving both the quality of life and the survival of ALS patients, in particular in patients without major bulbar muscles weakness [4,5]. There is a number of non-evidence based guidelines to support clinicians in their decision to start NIV in ALS patients. These take into account the clinical observation and the results of respiratory tests, and are based on consensus agreement [4]. However, no criteria is available to indicate the probability of respiratory failure within a defined time interval. In fact, in clinical practice, the decision to start NIV is highly dependent on the clinician's experience, together with NIV acceptance by the patient and caregivers. In this scenario, being able to predict the probability of a particular patient to progress to respiratory insufficiency within a certain period of time (such as before the next visit), would be of great clinical value. This information would have critical implications, regarding prognosis, health-costs and quality of life [6].

Unlike in other diseases such as cancer, cardiovascular diseases, Alzheimer's disease and Parkinson's disease [7,8], the state of the art in ALS relies on population-based approaches such as Kaplan–Meier survival tables and multivariable Cox proportional hazard

* Corresponding author.
E-mail addresses: andre.carreiro@tecnico.ulisboa.pt (A.V. Carreiro), sara.madeira@tecnico.ulisboa.pt (S.C. Madeira).

http://dx.doi.org/10.1016/j.jbi.2015.09.021
1532-0464/

regression models [9]. ALS studies have been tackling two major problems. The first is related to patients' diagnosis, studying the impact of diagnostic delay [10], the heterogeneity in ALS subtypes [11], or the diagnostic relevance of certain clinical features, such as axial muscles weakness [12,13]. The second concerns prognostic predictors, usually associated to ALS survival. The most studied prognostic factors are respiratory tests, such as the forced vital capacity (FVC) [14–18] and the maximal inspiratory and expiratory pressures (MIP/MEP) [14]. Some clinical features have also been identified as critical for prognosis, such as the site of onset (bulbar onset is generally associated with worse prognosis), weight loss and disease duration at diagnosis [17,19,20], the functional decline as assessed by the ALS Functional Rating Scale (ALSFRS) [16–18], muscle strength [15], age [17,19,20] and, possibly, gender [19,20]. The prognostic value of recent tests such as the phrenic nerve motor response [21], the respiratory subscore of the ALSFRS [16], as well as other respiratory tests [16,22], have also been explored.

In this scenario, this work proposes a new prognostic prediction approach allowing to answer a very important clinical question: "Given the patient's current condition (patient snapshot), will he/she be in respiratory insufficiency (RI) after a given period of time (time window)?". Our contribution is thus an innovative prognostic model able to evaluate the patient's current condition and, according to it, infer whether this specific patient will/will not require NIV in a given time window.

We first create patient snapshots using a new strategy based on hierarchical clustering with constraints, outperforming the current method based on pivot dates. We then compute learning examples based on the chosen time windows and use them to build classification models able to predict progression to assisted ventilation. To evaluate the proposed prognostic models, we use clinical data containing respiratory tests and neurophysiological data for 517 ALS patients, followed in the ALS clinic of the Translational Clinical Physiology Unit, Hospital de Santa Maria, Lisbon, during a period over 10 years. Since the medical appointments typically occur with a 90 days time interval, we build models for predicting the need for NIV for three time windows: (a) 90 days, the next medical appointment (short term); (b) 180 days, spanning two medical appointments from the current time (medium term), and (c) 365 days (long term). Promising results, as shown by an area under the receiver operating characteristics curve (AUC) value of approximately 79% for the three time windows, highlight the potential for such prognostic models to predict disease progression in clinical practice.

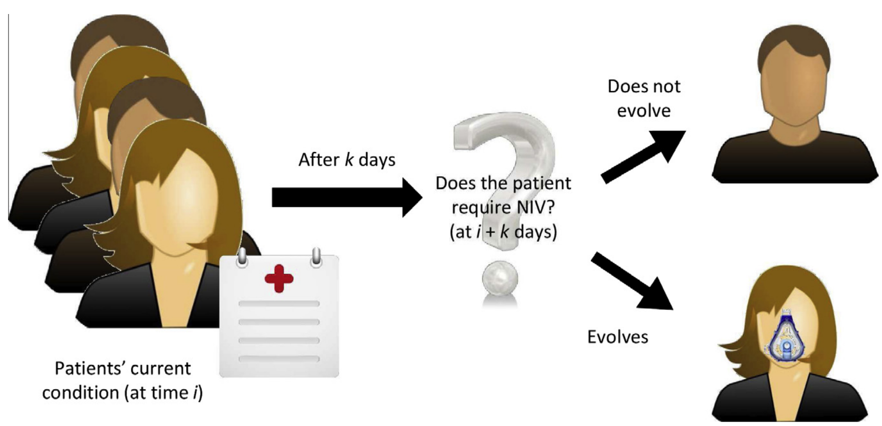

Fig. 1. Problem formulation: can we predict if a given patient will require non-invasive ventilation (NIV) after k days, using the patient's current condition?

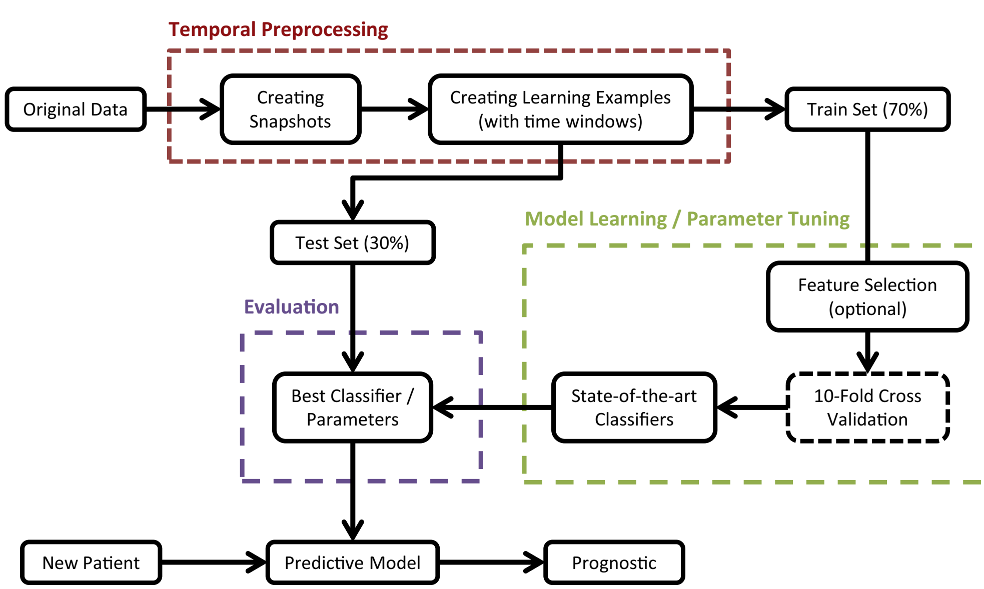

Fig. 2. Workflow of the proposed methodology for ALS prognostic prediction. The model is built based on a stratified 5 × 10-fold cross validation (CV) scheme (5 repetitions with different random seeds) using the training set (70% of patient snapshots). The performance is assessed using the test set (remaining 30%). The train-test stratified partition, as well as the 10 folds for CV, were constructed in order to guarantee that no patient has snapshots in both train and test sets, or in different CV folds.