Using Patient Information for the Prediction of Caregiver Burden in Amyotrophic Lateral Sclerosis

Anna Markella Antoniadi1,2, Miriam Galvin3, Mark Heverin3, Orla Hardiman1,4, Catherine Mooney1,2

1 FutureNeuro, the SFI Research Centre for Chronic and Rare Neurological Diseases; 2 School of Computer Science, University College Dublin, Belfield, Dublin 4, Ireland; 3 Academic Unit of Neurology, Trinity Biomedical Sciences Institute, Trinity College Dublin, Dublin 2, Ireland; 4 Department of Neurology, National Neuroscience Centre, Beaumont Hospital, Dublin 9, Ireland.

1. Introduction

Amyotrophic Lateral Sclerosis (ALS) is a fatal and currently incurable neurodegenerative disease that leads to the progressive death of motor neurons. Patients need support from at least one informal caregiver (usually a close relative). The inevitable and progressive nature of the disease can prove very challenging for the patients and their caregivers. Caregiver burden is a common issue that needs to be addressed; the term summarises the decline of the caregiver’s emotional or physical health, social life and financial status[1].

Our work has focused on using technology in the form of Clinical Decision Support Systems (CDSS), to assist with the timely prediction of a caregiver’s feelings of burden. We have previously created a model using 76 parameters (the patient’s and caregiver’s demographic, financial, and health status)[2], which would be possible to collect at each visit or to draw from the existing patient records.

Despite that model’s good performance, in this work we aim to reduce this number of features that the system uses in order to increase its usability and its compliance with the General Data Protection Regulations’ data minimisation principle. A review of papers on the factors associated with caregiver burden found evidence of correlation between caregiver burden and the patient’s physical functioning and behavioural impairment [3]. At this stage we wanted to discover how accurately we can predict caregiver burden using these features alone.

2. Data

Source
- Irish ALS Register and questionnaires on patients and their primary caregivers (demographics, socio-economic, resource use, QoL, anxiety and depression, burden) collected at 3 time-points
- 90 patient-caregiver dyads at baseline

Size
- 177 entries, 28 features

Features
- the patient’s sex, age and site of onset
- first symptom
- El Escorial diagnosis
- the ALS Functional Rating Scale-Revised (ALSFRS-R)


3. Methods

- Imputation method for missing data: Random Forest
- The prediction algorithm used is a Random Forest
- Data split in training and test set
- Trained and tested in 10-fold cross-validation (CV) using only the training set.
- Matthews correlation coefficient (MCC), ROC (Receiver Operator Characteristic) Curve, sensitivity and specificity to measure the quality of classifications.

Table 1. Distribution of caregivers into burden classes.

<table>
<thead>
<tr>
<th>Data</th>
<th>0 (Low Burden)</th>
<th>1 (High Burden)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full</td>
<td>74</td>
<td>103</td>
</tr>
<tr>
<td>Training</td>
<td>56</td>
<td>78</td>
</tr>
<tr>
<td>Testing</td>
<td>18</td>
<td>25</td>
</tr>
</tbody>
</table>

4. Results

The metrics for the independent test set and for the CV are shown in the table.

<table>
<thead>
<tr>
<th></th>
<th>MCC</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CV (Average)</td>
<td>0.33</td>
<td>0.67</td>
<td>0.66</td>
</tr>
<tr>
<td>Independent Test Dataset</td>
<td>0.47</td>
<td>0.84</td>
<td>0.61</td>
</tr>
</tbody>
</table>

The specificity of the model on the Independent Test Dataset is similar to that on the CV but the sensitivity is significantly improved when we ensemble the 10 individual models trained at each individual fold.

The model’s features in order of importance according to the mean decrease of the Gini Index are presented in the Figure to the right.

5. Conclusion

The four factors that were identified by the model as the most important according to the mean decrease of the Gini Index are: the patients’ age and age of disease onset, as well as the score that shows their level of ability to eat when they have and haven’t undergone a gastrostomy.

However, the results suggest that we can more accurately predict caregivers at risk of experiencing high burden while caring for a patient with ALS, than those that may experience low levels of burden. It is clear that the patient’s physical and cognitive impairment are not the only predictors of caregiver burden and a more accurate CDSS would require additional predictive information.

5. References