Reviewer's report

Title: Using machine learning algorithms to guide rehabilitation planning for home care clients

Version: 1 Date: 12 September 2007

Reviewer: Enzo Grossi

Reviewer's report:

General
The topic of the article is relevant, being the optimisation of rehabilitation programs one of unsolved problem in the home care setting.
Machine learning is an interesting technique and will certainly raise readers attention.
The message of the paper is clear. However there are some revision that the authors must respond before a decision on publication can be reached.

---------------------------------------------

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

Data set description.
The focus of the paper are the variables collected in the interRAI/minimum data set. The variables at the end belong to human subjects of whom no mention is made.The paper is based on huge data set concerning more than 24000 patients. Some descriptive statistics of this population according to the 19 variables collected could help the reader in putting the results in the appropriate clinical context.
The correlation index values between input and target variables should be described, to better circumstantiate the choice of a linear classifiers and the apparent non superiority of non linear classifiers.

Methods and Results.
Since the contribution offered by the article is mainly methodological and is focusing the strength of a specific statistical approaches, the description of the validation protocol is a fundamental procedure to verify the models' ability to generalize the results reached in the real world.
It is not clear from the description if the confusion matrix data described are just the result of interpolation or describe validation data. In this case we need to know: which protocol has been adopted like formal cross validation training-testing protocol( like so-called 5x2 cross-validation protocol described by Dietterich in 1998), resampling methods such as the bootstrap, or K-fold, in which we subdivide the data into K roughly equal-sized parts, then repeat the modeling process K times, leaving one section out each time for validation
purposes.
In addition the target variable should be precisely defined (mobility in bed?).
In the tables 2, 4, 5, 7 the overall accuracy rate should be added to false+ and
false- rates to facilitate the comparison among different techniques.

-------------------------------------------------------------------------

Minor Essential Revisions (such as missing labels on figures, or the wrong use of
a term, which the author can be trusted to correct)
Title page: the title could be more direct as for example: role of machine learning
algorithms in optimisation of rehabilitation protocols. The e-mail addresses for all
authors must be included on the title page
Authors' contribution: I suggest to add the following sentence: “All authors read
and approved the final manuscript”
Figure Legends: the legends should be included in the main text file immediately
following the references, rather than being a part of the figure file

-------------------------------------------------------------------------

Discretionary Revisions (which the author can choose to ignore)
Style and language: Even if there is no explicit limit on the length of articles
submitted, authors are encouraged to be concise. The style of the article is rather
narrative. I recommend to be
more short and snappy.

What next?: Unable to decide on acceptance or rejection until the authors have
responded to the major compulsory revisions

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:
'I declare that I have no competing interests'