Predicting Motor and Cognitive Improvement Through Machine Learning Algorithm in Human Subject that Underwent a Rehabilitation Treatment in the Early Stage of Stroke

Patrizio Sale, *,§[1] Giorgio Ferriero, †,[1] Lucio Ciabattoni, ‡,[1] Anna Maria Cortese, § Francesco Ferracuti, ‡ Luca Romeo, ‡ Francesco Piccione, § and Stefano Masiero, *

Background: The objective of this study was to investigate, in subject with stroke, the exact role as prognostic factor of common inflammatory biomarkers and other markers in predicting motor and/or cognitive improvement after rehabilitation treatment from early stage of stroke. *Methods:* In this longitudinal cohort study on stroke patients undergoing inpatient rehabilitation, data from 55 participants were analyzed.

Functional and clinical data were collected after admission to the rehabilitation unit. Biochemical and hematological parameters were obtained from peripheral venous blood samples on all individuals who participated in the study within 24 hours from the admission at the rehabilitative treatment. Data regarding the health status were collected at the end of rehabilitative treatment.

First, a feature selection has been performed to estimate the mutual dependence between input and output variables. More specifically, the so called Mutual Information criterion has been exploited.

In the second stage of the analysis, the Support Vector Machines (SVMs), a non-probabilistic binary machine learning algorithm widely used for classification and regression, has been used to predict the output of the rehabilitation process.

Performances of the linear SVM regression algorithm have been evaluated considering a different number of input features (ranging from 4 to 14). The performance evaluation of the model proposed has been investigated in terms of correlation, Root Mean Square Error (RMSE) and Mean Absolute Deviation Percentage (MADP). *Results:* Results on the test samples show a good correlation between all the predicted and measured outputs (i.e. T1 Barthel Index (BI), T1 Motor Functional Independence Measure (FIM), T1 Cognitive FIM and T1 Total FIM) ranging from 0.75 to 0.81. While the MADP is high (i.e., 83.96%) for T1 BI, the other predicted responses (i.e., T1 Motor FIM, T1 Cognitive FIM, T1 Total FIM) disclose a smaller MADP of 30%. Accordingly, the RMSE ranges from 4.28 for T1 Cognitive FIM to 22.6 for T1 BI. *Conclusions:* In conclusion, the authors developed a new predictive model using SVM regression starting from common inflammatory biomarkers and other ratio markers. The main efforts of our model have been accomplished in regard to the evidence that the type of stroke has not shown itself to be a critical

From the *Rehabilitation Unit, Department of Neuroscience, University of Padua, Padua, Italy; †Department of Physical Medicine and Rehabilitation, Scientific Institute of Lissone, IRCCS, Istituti Clinici Scientifici Maugeri, Lissone MB, Italy; ‡Department of Information Engineering, Università Politecnica delle Marche, Ancona, Italy; and §San Camillo Hospital IRCCS, Venice, Italy.

Received January 22, 2018; revision received May 18, 2018; accepted June 17, 2018.

Address correspondence to Patrizio Sale, Rehabilitation Unit, Department of Neuroscience, University of Padua, Via Giustiniani 2, Padua 35128, Italy E-mail: patrizio.sale@gmail.com giorgio.ferriero@fsm.it l.ciabattoni@univpm.it annamaria.cortese@ospedalesancamillo.net f.ferracuti@univpm.it l.romeo@univpm.it francesco.piccione@ospedalesancamillo.net stef.masiero@unipd.it

[1]These authors contributed equally to this work.

1052-3057/$ - see front matter

https://doi.org/10.1016/j.jstrokecerebrovasdis.2018.06.021

input variable to predict the discharge data, furthermore, among the four selected indicators, Barthel at T1 is the less predictable (MADP > 80%), while it is possible to predict T1 Cognitive FIM with an MADP less than 18%.

Keywords: Stroke—Inflammation—Predictive values—Rehabilitation—SVM regression and FIM

Background

Stroke is a leading cause of serious, long-term disability, in adult population and it is the second leading cause of death for older people in high-income countries.[1] Deficits may include motor and cognitive impairments such as partial paralysis, difficulties with memory, thinking, language, and movements. The pathophysiology of cerebral ischemia has played a key role in guiding biomarker research in ischemic stroke.[2] Mounting evidence suggests that immunity and inflammation are key elements involved in several steps of ischemic cascade such as acute events after blood flow reduction, brain tissue damage, progression of ischemic lesions, and tissue repair.[3,4] Neutrophil/granulocytes (neutrophils)[5] and macrophages/microglia[6] have been showed as the main players in postischemic infarction. Moreover, have been demonstrated that T-cells also have an impact on tissue damage.[6] In the last years, there have been many efforts to find useful diagnostic markers to monitor stroke patients. Serum biomarkers are considered to be the most valuable adjunct to routine clinical examination and imaging data.[7] As reported by Hudack and colleagues, prognostic information after neurological disease comprises 4 characteristics of disorders: response, remission, recurrence, and duration[8] but the knowledge in this area is still insufficient. Studying the dynamics of inflammatory biomarkers during rehabilitation could be worthwhile to better understand the mechanism and also to innovate the treatment protocols.

The aim of this study is to understand, in subjects with stroke that underwent rehabilitative training, the exact role as prognostic factor of most common inflammatory biomarkers in prediction of motor or cognitive improvement after rehabilitation treatment from the early stage of stroke. In this context, different works propose classification and regression models for the prediction of functional outcomes in ischemic stroke patients by using machine learning algorithms[9-11] such as support vector machines, which can be applied efficiently to feature spaces whose dimensionality exceeds the number of data points, including infinite feature spaces.

Methods

The inclusion criteria are: sub-acute stroke patients (15 ± 10 days from injury), unilateral paresis. The following demographic and clinical data were collected within 24 hours from the admission at the rehabilitation unit (T0): age (number), sex (male or female), type of stroke (ischemic large vessel/hemorrhagic), first event (yes/no), recombinant tissue plasminogen activator (rTPA) (yes/no)[12], Barthel Index (BI)[13], functional independence measure (FIM).[13]

At discharge (T1) the following data were collected: motor impairment (yes/no), dysphagia (yes/no), tracheostomy (yes/no), neuropsychological impairment (yes/no), speech impairment (yes/no), presence of nasogastric feeding tube (yes/no), length of stay (rLOS)[14], BI, FIM.

The FIM is a basic indicator of severity of disability for evaluating activities of daily living (ADL) in subjects with stroke and measures the level of a patient's disability indicating how much assistance is required for the individual to carry out ADL. Total scores range from 18 to 126 and indicate to what extent an individual is capable to reach independence in ADL. In particular, FIM contains 18 items composed of: 13 motor tasks (Motor FIM) (from minimum of 13 and maximum of 91) and 5 cognitive tasks (considered basic ADL) (Cognitive FIM) (from minimum of 5 and maximum of 35). Tasks are rated on a 7-point ordinal scale that ranges from total assistance (or complete dependence) to complete independence. We used the FIM effectiveness (eFIM), calculated as (FIM at discharge−FIM on admission)/(126−FIM on admission) according to Koh and colleagues.[15] eFIM, which is independent from FIM on admission, has been used to compare patients with various degrees of disability severity. The rLOS value represents the total number of rehabilitation days from admission to discharge.[14]

Biochemical and Hematological Parameters: we obtained peripheral venous blood samples of all individuals who participated to the study within 24 hours from the admission at the rehabilitative treatment (T0) to investigate complete blood count and inflammatory profile [White Blood Cells; Fibrinogen; Anti-thrombin; D-Dimer; C-reactive protein (RCP); erythrocyte sedimentation rate (VES); Erythrocytes; Hemoglobin; Hematocrits; Hemoglobin distribution width (RDW); Platelets; Neutrophil; Lymphocytes; Monocytes; Eosinophils; Basophils; Lymphocytes CD3; Lymphocytes CD4; Lymphocytes CD3+ CD8+; Ratio Lymphocytes CD4/CD8; Lymphocytes Natural Killer (NK); Lymphocytes CD19]. Neutrophil-to-lymphocyte ratio (NLR) is calculated by dividing the number of neutrophils by number of lymphocytes.[16-23] Platelets-lymphocyte ratio is calculated by dividing the number of platelets by number of lymphocytes.

Each patient was informed about the study procedure and aims. Then, after a period of discussion and reflection,