Ann Ig 2022 Jan-Feb; 34(1): 79-83. doi: 10.7416/ai.2021.2444. Epub 2021 Apr 2

SHORT PAPER

Application of machine learning techniques to physical and rehabilitative medicine

V. Santilli[1,2]

Key words: Machine learning, rehabilitation, data mining, disability, big data
Parole chiave: Machine learning, riabilitazione, data mining, disabilità, big data

Abstract

Nowadays, digital information has increased exponentially in every field to such an extent that it generates huge amounts of electronic data, namely Big Data. In the field of Artificial Intelligence, Machine Learning can be exploited in order to transform the large amount of information to improve decision-making. We retrospectively evaluated the data collected from 2016 to 2018, using the database of approximately 4000 rehabilitation hospital discharges (SDO) of the Latium Region (Italy). Three models of machine learning algorithms were considered: Support of vector machine; Neural networks; Random forests. Applying this model, the estimate of the average error is 9.077, and specifically, considering the distinction between orthopedic and neurological patients, the average error obtained is 7.65 for orthopedic and 10.73 for neurological patients. SDO information flow can be used to represent and quantify the potential inadequacy and inefficiency of rehabilitation hospitalizations, although there are limitations such as the absence of description of pre-pathological conditions, changes in health status from the beginning to the end of hospitalization, specific short- and long-term outcomes of rehabilitation, services provided during hospitalization, as well as psycho-social variables. Furthermore, information from wearable devices capable of providing clinical parameters and movement data could be integrated into the dataset.

Nowadays, digital information has increased exponentially in every field to such an extent that it generates huge amounts of electronic data, namely *Big Data*. Specifically, the healthcare sector is one of the most important players of this phenomenon (1), especially by virtue of diagnostic imaging digitalization (2), as well as the growing development and use of the Informatrion Technology tools (IT tools), such as electronic medical records for collection of patient data (3). In response to the increasing need for a wide source of information, the computerized medical

[1] *Digital Med, Start-Up Sapienza University of Rome, Italy*
[2] *Department of Anatomical and Histological Sciences, Legal Medicine and Orthopedics, Sapienza University of Rome, Italy*

Annali di Igiene : Medicina Preventiva e di Comunità (Ann Ig)
ISSN 1120-9135 https://www.annali-igiene.it

record plays a central role in allowing documentary traceability of the various activities carried out, but also in allowing easy access to data for clinical-scientific, administrative, managerial research and medico-legal purposes.

Big Data is characterized by five parameters: Volume, Velocity, Variety, Veracity and Value (4). In the field of Artificial Intelligence, machine learning techniques can be exploited in order to transform the large amount of information to improve decision-making, to increase effectiveness and efficiency and reduce human errors and costs, especially in the healthcare sector (5). In Machine Learning, a model learns from the available examples by defining the relationship between input and output data and generating predictive data.

Within the healthcare sector, and in our specific field of rehabilitation assistance, the need to organize rehabilitation interventions, and care facilities meeting criteria of adequacy, is growing. Effectiveness and efficiency are also fundamental elements for medical rehabilitation and are in particular associated with intervention adequacy. Effectiveness (RES) and Efficiency (REY) are indices (6) related to the care processes improvement, to the choice of the level and type of rehabilitation, as well as to health expenditure, and are important in defining the classification algorithm.

All these considerations have directed a part of my research activity in the field of artificial intelligence applications to physical and rehabilitative medicine and, in 2016, I founded a university start-up of the Sapienza University of Rome, <Digital Med srl>, which has the data science among its research and development areas.

Digital Med collaborated from 2018 till today to a regional project that led to the development of a computerized medical record with an integrated machine learning component.

The evaluation of the functional status described by the patient's degree of autonomy exploited the Barthel Index (BI) score (7) and the difference between admission and discharge BI values was correlated to the effectiveness of rehabilitation intervention.

We retrospectively evaluated the data collected from 2016 to 2018, using the database of approximately 4,000 hospital discharges (in Italy known as the Hospital Discharge Card, SDO) of the Latium Region (Italy). The following patient inclusion criteria were applied: age ≥18 years; time between the onset of the disease and rehabilitation hospitalization ≤60 days; length of hospitalization >14 and ≤90 days; first hospitalization. Patients with missing data (e.g. on admission or discharge functional scores, or other data regarding the onset of disabling conditions) were excluded (8).

Specifically, starting from the data collected at patient's admission, the algorithm was able to predict (to classify) the value of the discharge BI, or to predict rehabilitation intervention effectiveness with a certain margin of error.

Three models of machine learning algorithms were considered: 1) Support of vector machine; 2) Neural networks; 3) Random forests.

The three models were trained for each type of patient (orthopedic and neurological). Once the prediction of the three models has been obtained, the value of the BI is equal to the average of the three values obtained.

Applying this model, the estimate of the average error is 9.077, and specifically, considering the distinction between orthopedic and neurological patients, the average error obtained is 7.65 for orthopedic and 10.73 for neurological patients. It should be remembered that these scores have their limits and none of these can confidently predict the real functional evolution of the patient in terms of autonomy.

SDO information flow can be used to represent and quantify the potential inadequacy and inefficiency of rehabilitation

hospitalizations, although there are limitations, such as the absence of description of pre-pathological conditions, changes in health status from the beginning to the end of hospitalization, specific short- and long-term outcomes of rehabilitation, services provided during hospitalization, as well as psycho-social variables. It may be useful to estimate the risk of complications of a patient at the beginning of hospitalization, in order to better manage health resources, anticipating the needs that the patient may present, in order to plan an adequate transfer to the most appropriate unit.

The objective of Digital Med is to extend the potential of this tool to all information and data present in the electronic medical records, and in this sense, we are trying to establish collaborative contacts with important rehabilitation structures on the national territory. The goal is to create a "Data Lake" of health data on which to train algorithms.

Since 2019 I have been working on the same dataset (8) in order to improve the previously observed results, also including days of hospitalization (LD) as a predictable variable. The latter, together with the BI forecast, could also provide information on the rehabilitation intervention efficiency as well as effectiveness (9, 10). LD can be defined as an appropriate length of time for each type of rehabilitation and for each healthcare setting. The objective could be to define: a minimum limit of days of hospitalization, below which an area of potential organizational inadequacy is determined, that is presumably related to healthcare setting type where hospitalization is provided; and a maximum limit of days of hospitalization, above which an area of potential inefficiency is detected, generated by a possible excess of hospitalization time.

The following models were used: xGBT (Extreme Gradient Boosting Tree), a powerful ensemble model fast and accurate for classification and regression problems; lightGBM (Light Gradient Boosting Machine), also a powerful tool created by Microsoft, a leading algorithm in many tasks in machine learning; Catboost, also a fast and efficient algorithm that achieves state of the art results on similar tasks. The results obtained are shown in tables 1 and 2.

An open question could still be whether the information on functional outcomes from BI alone can improve effectiveness and efficiency of rehabilitation processes and set the right indications to guide the corrective processes, in consideration of other variables that influence the treatment processes, such as pre-pathological condition, initial clinical complexity, treatment setting, psychosocial factors, possible complications that occurred during management (11).

Table 1 - Comparison of the results previously obtained with the new models applied for the prediction of the Barthel Index with orthopedic patients

Model	Error rate in 'Orthopedic' (RMSE)
xGBT	6.71
lightGBM	**6.54**
Catboost	6.8
Our previous work	**7.65**
Our Customized model	6.58

Legend: xGBT: Extreme Gradient Boosting tree; lightGBM: Light Gradient Boosting Machine.
RMSE: The Root mean squared error (RMSE) of an estimator measures the average of the squares of the errors that is, the average squared difference between the estimated values and the actual value.

Table 2 - Comparison of the results previously obtained with the new models applied for the prediction of the Barthel Index with neurologic patients.

Model	Error rate in 'Neurologic' (RMSE)
xGBT	8.9
lightGBM	9.23
Catboost	9.09
Our previous work	**10.73**
Our Customised model	**8.66**

Legend: xGBT: Extreme Gradient Boosting tree; lightGBM: Light Gradient Boosting Machine.
RMSE: The Root mean squared error (RMSE) of an estimator measures the average of the squares of the errors that is, the average squared difference between the estimated values and the actual value.

Conclusions

In the light of what has been elaborated up to this point, and in consideration of the possible criticalities highlighted, Digital Med proposes to work with machine learning techniques operating on data sets that contain a variety of information, not only attributable to elements of a purely clinical nature, but also related to family, lifestyle, social and environmental spheres, etc. Furthermore, information from wearable devices, capable of providing clinical parameters and movement data, could be integrated into the dataset.

Disclose of interest: The author reports no conflict of interest.
Funding source: None
Disclosure of relationships and activities: The author reports no relationships and activities.

Riassunto

Applicazione delle tecniche di machine learning alla medicina fisica e riabilitativa

Al giorno d'oggi, le informazioni digitali sono aumentate in modo esponenziale in ogni campo a tal punto da generare enormi quantità di dati elettronici, ovvero i Big Data. Nel campo dell'Intelligenza Artificiale, il Machine Learning può essere sfruttato per trasformare la grande quantità di informazioni per migliorare il processo decisionale. Abbiamo valutato retrospettivamente i dati raccolti dal 2016 al 2018, utilizzando il database di circa 4000 dimissioni ospedaliere di riabilitazione (SDO) della Regione Lazio (Italia). Sono stati considerati tre modelli di algoritmi di machine learning: 1) Support of vector machine; 2) Neural networks; 3) Random forests. Applicando questo modello, la stima dell'errore medio è di 9,077, e nello specifico, considerando la distinzione tra pazienti ortopedici e neurologici, l'errore medio ottenuto è 7,65 per i pazienti ortopedici e 10,73 per i pazienti neurologici. Il flusso di informazioni SDO può essere utilizzato per rappresentare e quantificare la potenziale inadeguatezza e inefficienza dei ricoveri riabilitativi, sebbene vi siano limitazioni come l'assenza di descrizione delle condizioni pre-patologiche, i cambiamenti dello stato di salute dall'inizio alla fine del ricovero, breve specifico - e gli esiti a lungo termine della riabilitazione, i servizi forniti durante il ricovero, nonché le variabili psicosociali. Inoltre, le informazioni provenienti da dispositivi indossabili in grado di fornire parametri clinici e dati di movimento potrebbero essere integrate nel set di dati.

References

1. MacEachern SJ, Forkert ND. Machine Learning for Precision Medicine. Genome 2020 Oct 22. doi: 10.1139/gen-2020-0131.
2. Kapoor N, Lacson R, Khorasani R. Workflow Applications of Artificial Intelligence in Radiology and an Overview of Available Tools. J Am Coll Radiol 2020 Nov; **17**(11): 1363-70. doi: 10.1016/j.jacr.2020.08.016.
3. Beam AL, Kohane IS. Big Data and Machine Learning in Health Care. JAMA 2018 Apr 3; **319**(13): 1317-8. doi: 10.1001/jama.2017.18391.

4. Phinyomark A, Petri G, Ibáñez-Marcelo E, Osis ST, Ferber R. Analysis of Big Data in Gait Biomechanics: Current Trends and Future Directions. J Med Biol Eng 2018; **38**(2): 244-60. doi: 10.1007/s40846-017-0297-2. Epub 2017 Jul 17.
5. Bates DW, Saria S, Ohno-Machado L, Shah A, Escobar G. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. Health Aff (Millwood) 2014 Jul; **33**(7): 1123-31. doi: 10.1377/hlthaff.2014.0041.
6. Damiani C, Mangone M, Paoloni M, et al. Trade-Offs with rehabilitation Effectiveness (REs) and Efficiency (REy) in a sample of Italian disabled persons in a in post-acuity rehabilitation unit. Ann Ig 2020 Jul-Aug; **32**(4): 327-35. doi: 10.7416/ai.2020.2356.
7. Seccia R, Boresta M, Fusco F, et al. Data of patients undergoing rehabilitation programs. Data Brief 2020 Mar 16; **30**: 105419. doi: 10.1016/j.dib.2020.105419.
8. Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. Int Disabil Stud 1988; **10**(2): 61-3. doi: 10.3109/09638288809164103.
9. Krause DD. Data Lakes and Data Visualization: An Innovative Approach to Address the Challenges of Access to Health Care in Mississippi. Online J Public Health Inform 2015 Dec 30; 7(3): e225. doi: 10.5210/ojphi.v7i3.6047.
10. Chow P, Chen C, Cheong A, et al. Factors and trade-offs with rehabilitation effectiveness and efficiency in newly disabled older persons. Arch Phys Med Rehabil 2014 Aug; **95**(8): 1510-20e4. Epub 2014 Apr 12.
11. Caldas R, Mundt M, Potthast W, Buarque de Lima Neto F, Markert B. A systematic review of gait analysis methods based on inertial sensors and adaptive algorithms. Gait Posture 2017 Sep; **57**: 204-10. doi: 10.1016/j.gaitpost.2017.06.019. Epub 2017 Jun 24.

Corresponding Author: Valter Santilli, MD, Department of Anatomical and Histological Sciences, Legal Medicine and Orthopedics, Sapienza University, Piazzale Aldo Moro 5, 00185, Rome, Italy
e-mail: valter.santilli@uniroma1.it