



Borsa di studio attivata ai sensi di quanto disposto dal D.M. n. 1061 del 10/08/2021

Titolo del progetto: Tecnologie innovative basate sul data mining per supportare la diagnosi precoce delle malattie croniche

La borsa sarà attivata sul seguente corso di dottorato accreditato per il XXXVII ciclo:

AUTOMATICA, BIOINGEGNERIA E RICERCA OPERATIVA - ABRO

Responsabile scientifico: Laura Palagi

Area per la quale si presenta la richiesta: INNOVAZIONE

Numero di mensilità da svolgere in azienda: 6

Numero di mensilità da svolgere all'estero: 6 presso Da definire

Azienda: Spindox Labs Srl

Il Dipartimento è disponibile a cofinanziare per un importo pari a euro: 10000

Dipartimento finanziatore: DIPARTIMENTO DI INGEGNERIA INFORMATICA, AUTOMATICA E GESTIONALE - ANTONIO RUBERTI- con delibera del 20/09/2021

Progetto di ricerca:

La previsione di insorgenza e decorso delle patologie multifattoriali croniche è un problema complesso, visto che entrambe sono influenzati da molti fattori non modificabili (essenzialmente, genetici) e modificabili (principalmente ambientali, lavorativi e in generale legati allo stile di vita), ciascuno con basso peso. Ne consegue che il decorso della malattia è estremamente variabile da soggetto a soggetto e non è possibile prevederlo in modo affidabile.

Tra queste patologie, molto paradigmatica appare la Sclerosi Multipla (SM), malattia autoimmune del Sistema Nervoso Centrale, che rappresenta la seconda causa di invalidità neurologica nei giovani adulti, dopo i traumi. La sclerosi multipla (SM) è caratterizzata da un decorso estremamente variabile, che va da forme benigne a altamente invalidanti. Di solito, essa inizia con una fase remittente-ricidivante (RR-MS), che poi procede a una forma secondariamente progressiva (SP-SM), con grave accumulo disabilità e scarsa o nulla risposta alle terapie disponibili. I trattamenti modificanti il decorso offrono la possibilità di adattare la terapia alle esigenze di ciascun paziente, ma mancano indicatori prognostici affidabili e modelli convalidati in grado di prevedere il decorso della SM. Poiché la progressione della malattia è accompagnata da accumulo di disabilità, l'identificazione precoce dei pazienti a forte rischio di progressione consentirebbe di iniziare la somministrazione di terapie ad alto impatto (che hanno possibili effetti collaterali importanti) prima che si sviluppino danni irreversibili.

Modelli prognostici affidabili sarebbero di grande utilità clinica, fornendo supporto nelle scelte terapeutiche, con una significativa riduzione dei costi economici e sociali della malattia. Dato il fallimento dei tradizionali metodi di previsione, approcci innovativi basati sull'apprendimento automatico (ML) sono ampiamente testati [2]. Gli algoritmi ML sono in grado di catturare relazioni complesse tra i dati meglio dei modelli di regressione standard. Risultati promettenti sono stati ottenuti per diverse patologie neurologiche, mentre nel campo di patologie multifattoriali croniche, quale la SM, gli strumenti disponibili hanno bisogno di miglioramenti sostanziali per essere applicabili nella pratica clinica. Utilizzando una varietà di approcci ML sul database relativamente piccolo di un singolo centro, è stato recentemente dimostrato [1] che dati della routine clinica possono essere utilizzati per prevedere il decorso della SM in singoli pazienti, predicendo se passeranno da RR-MS a SP-MS in 2 anni. I risultati ottenuti sono stati buoni, sebbene con un numero rilevante di previsioni false positive e su un intervallo di tempo (2 anni) di utilità clinica

limitata.

L'esistenza di un registro nazionale dei pazienti con SM, fornito da Federazione Italiana Sclerosi Multipla, consente di disporre di un database molto ampio.

Questo tipo di approccio potrebbe essere esteso ad altre patologie multifattoriali croniche che derivano da complesse interazioni fra fattori modificabili e non, quali ad esempio il diabete mellito di tipo 2 (DM2), una delle patologie che stanno aumentando più rapidamente a livello mondiale. DM2 aumenta fino a 2 volte il rischio di morte per eventi cardiovascolari ed ha molte conseguenze negative sulla salute delle persone colpite. Si stima che il costo annuo delle spese sanitarie per i pazienti con DM2 sia circa il doppio di quello per soggetti non diabetici. Per questa patologia si avrebbero a disposizione i dati forniti dal registro nazionale Associazione Italiana Medici Diabetologici (AMD).

L'associazione fra stile di vita e insorgenza/progressione di malattia è sicuramente molto alta per DM2, più da indagare per SM. Comunque, la maggior parte degli studi sono stati condotti chiedendo a piccoli gruppi di pazienti di modificare il loro stile di vita per brevi periodi e per lo più considerando solo parametri isolati (ad esempio dieta, allenamento fisico, integrazione di vitamina D).

Per studiare adeguatamente come stile di vita influenza il decorso di patologie croniche, si dovrebbe invertire l'approccio, monitorando (eventualmente utilizzando dispositivi indossabili e APP per smartphone) alcune caratteristiche della vita quotidiana dei pazienti e analizzare la loro relazione con il decorso clinico utilizzando Intelligenza artificiale

La presente proposta si propone di:

- definire modelli di ottimizzazione, data mining e machine learning per la predizione precoce del decorso della malattia;
- definire algoritmi predittivi del decorso della malattia nel singolo paziente, consentendo di limitare gli effetti collaterali della terapia in pazienti a basso rischio di progressione e di dirottare risorse su pazienti più a rischio;
- identificare le variabili di maggior interesse per la profilazione del paziente e la successiva impostazione terapeutica;
- Dare impulso alla creazione di registri di informazione sanitaria che contengano informazioni sullo stile di vita dei pazienti
- validare tali strumenti analitici sulla base dell'esperienza dei medici;
- contribuire alla ricerca clinica con possibili ricadute su controllo della disabilità e dei costi socio-economici.
- Fornire alle persone con patologie croniche uno strumento per migliorare la capacità di gestire la malattia, alimentando il controllo consapevole di quei fattori che possono migliorarne (ovvero aggravarne) il decorso, in un'ottica di "empowerment" dei pazienti.

[1] Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis, R Seccia, D Gammelli, F Dominici, S Romano, AC Landi, M Salvetti, et al., PLoS one 15 (3), e0230219, 2020

[2] Machine Learning Use for Prognostic Purposes in Multiple Sclerosis, R Seccia, S Romano, M Salvetti, A Crisanti, L Palagi, F Grassi, Life 11 (2), 1221, 2021

Titolo del progetto (inglese): Data mining-based innovative technologies to support early diagnosis of chronic diseases

Progetto di ricerca (inglese):

Predicting the onset and course of chronic multifactorial diseases is a complex problem, given that both are influenced by many non-modifiable (essentially, genetic) and modifiable (mainly environmental, occupational and generally lifestyle-related) factors, each with a low weight. It follows that the course of the disease is extremely variable from person to person and cannot be reliably predicted.

Among these pathologies, multiple sclerosis (MS), an autoimmune disease of the central nervous system, is very paradigmatic, representing the second most common cause of neurological disability in young adults, after trauma. Multiple sclerosis (MS) is characterised by an extremely variable course, ranging from benign to highly disabling forms. It usually starts with a remitting-recurrent phase (RR-MS), which then progresses to a secondarily progressive form (SP-MS), with severe disability accumulation and little or no response to available therapies. Course-modifying treatments offer the possibility to tailor therapy to the needs of each patient, but there is a lack of reliable prognostic indicators and validated models to predict the course of MS. As disease progression is accompanied by accumulation of disability, early identification of patients at high risk of progression would allow the initiation of high-impact therapies (which may have major side effects) before irreversible damage develops.

Reliable prognostic models would be of great clinical utility, providing support in treatment choices, with a significant reduction in the economic and social costs of the disease. Given the failure of traditional prediction methods, innovative machine learning (ML) approaches are widely tested [2]. ML algorithms are able to capture complex relationships between data better than standard regression models. Promising results have been obtained for several neurological diseases, whereas in the field of chronic multifactorial diseases, such as MS, the available tools need substantial improvements to be applicable in clinical practice. Using a variety of ML approaches on the relatively small database of a single centre, it has recently been demonstrated [1] that data from clinical routines can be used to predict the course of MS in individual patients by predicting whether they will switch from RR-MS to SP-MS in 2 years. The results obtained were good, although with a significant number of false positive predictions and over a time interval (2 years) of limited clinical utility.

The existence of a national register of MS patients, provided by the Italian Multiple Sclerosis Federation, provides a very large database.

This type of approach could be extended to other chronic multifactorial diseases resulting from complex interactions between modifiable and non-modifiable factors, such as type 2 diabetes mellitus (DM2), one of the fastest growing diseases worldwide. DM2 increases the risk of death from cardiovascular events by up to 2 times and has many negative health consequences for those affected. It is estimated that the annual cost of health care for patients with DM2 is about twice that of non-diabetic individuals. Data from the national register of the Italian Association of Diabetes Physicians (AMD) would be available for this disease.

The association between lifestyle and disease onset/progression is certainly very high for DM2, more to be investigated for MS. However, most studies have been conducted by asking small groups of patients to modify their lifestyle for short periods and mostly considering only isolated parameters (e.g. diet, physical training, vitamin D supplementation).

In order to properly study how lifestyle influences the course of chronic diseases, the approach should be reversed, by monitoring (possibly using wearable devices and smartphone APPs) some characteristics of patients' daily life and analysing their relationship with the clinical course using artificial intelligence.

This proposal aims at

- design optimisation, data mining and machine learning models for the early prediction of the disease course;
- define predictive algorithms of the disease course in individual patients, allowing to limit the side effects of therapy in patients at low risk of progression and to divert resources to patients at higher risk;
- identify the variables of greatest interest for patient profiling and the subsequent therapeutic approach;
- boosting the creation of health information registers containing information on patients' lifestyles
- validate these analytical tools on the basis of doctors' experience;
- contribute to clinical research with possible spin-offs on disability control and socio-economic costs.
- To provide people with chronic diseases with a tool to improve their ability to manage the disease, by nurturing

conscious control of those factors that can improve (or worsen) its course, with a view to patient empowerment.

[1] Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis, R Seccia, D Gammelli, F Dominici, S Romano, AC Landi, M Salvetti, et al., PloS one 15 (3), e0230219, 2020

[2] Machine Learning Use for Prognostic Purposes in Multiple Sclerosis, R Seccia, S Romano, M Salvetti, A Crisanti, L Palagi, F Grassi, Life 11 (2), 1221, 2021