



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
CORSO DI LAUREA MAGISTRALE IN BIOINGEGNERIA

PREDIZIONE DELLA PROGRESSIONE DI SCLEROSI MULTIPLA MEDIANTE TECNICHE DI MACHINE LEARNING

Relatore

Prof.ssa Martina VETTORETTI

Correlatore

Dott.ssa Erica TAVAZZI

Candidato

Davide ALBERTINI

1242573

Anno Accademico 2021-2022

Indice

1	Sclerosi Multipla e modelli predittivi	3
1.1	Premessa	3
1.2	Sclerosi Multipla	4
1.2.1	Eziologia	4
1.2.2	Fisiopatologia	5
1.3	Sintomi	5
1.4	Diagnosi	6
1.5	Fenotipi	7
1.5.1	Clinically Isolated Syndrome (CIS)	8
1.5.2	Relapse-Remitting Multiple Sclerosis (RRMS)	8
1.5.3	Secondary Progressive Multiple Sclerosis (SPMS)	9
1.5.4	Primary Progressive Multiple Sclerosis (PPMS)	10
1.6	Trattamenti	11
1.6.1	Terapie di base	11
1.6.2	Terapie di escalation	13
1.7	Modelli predittivi per la Sclerosi Multipla	14
1.8	EDSS e metodi di classificazione	16
1.8.1	Modello secondo Zhao et al., 2017	17
1.8.2	Modello secondo Zhao et al. 2020	17
1.8.3	Modello secondo Rocca et al., 2021	18
1.8.4	Modello secondo Pinto et al., 2020	18
1.8.5	Modello secondo Law et al., 2019	19
1.8.6	Modello secondo Tommasin et al., 2021	19
1.8.7	Modello secondo Yperman et al., 2020	20
1.9	Scopi e organizzazione della tesi	21
2	Il database MSOAC Placebo: analisi e pre-processing	23
2.1	Il database MSOAC Placebo	23
2.2	Analisi e pre-processing	26

2.2.1	Dominio DM	27
2.2.2	Dominio FA	30
2.2.3	Dominio FT	30
2.2.4	Dominio QS	32
2.2.5	Dominio SC	34
2.2.6	Dominio OE	35
2.2.7	Dominio MH	36
2.3	Feature Selection by NAs e andamento temporale del parametro EDSS	37
3	Pipeline per lo sviluppo dei modelli predittivi	49
3.1	Pipeline riassuntiva per i modelli	49
3.2	Orizzonti temporali di previsione	53
3.3	Definizione di “Worsening”: le soglie	54
3.4	Split in Training set e Test set	54
3.5	Fattorizzazione delle variabili categoriche	55
3.6	Imputazione dei dati mancanti: MICE	56
3.7	Normalizzazione delle variabili	58
3.8	Regressione Logistica	59
3.9	Stepwise Feature Selection	59
3.10	Sbilanciamento tra classi	60
4	Risultati	63
4.1	Modelli per predizioni a 6 mesi	63
4.2	Modelli per predizioni a 12 mesi	66
4.3	Modelli per predizioni a 18 mesi	69
5	Conclusioni e sviluppi futuri	73

Introduzione

La Sclerosi Multipla è una malattia che colpisce il sistema nervoso centrale causando un danneggiamento transitorio o permanente della guaina mielinica che avvolge gli assoni delle cellule neuronali. Le conseguenze, per un paziente affetto da questa patologia, sono disabilità a livello motorio e cognitivo.

Al giorno d'oggi, sono identificate quattro forme di Sclerosi Multipla: Recidivante - Remittente, Primariamente progressiva, Secondariamente progressiva e Sindrome Clinicamente Isolata.

Attualmente non esiste una cura per la Sclerosi Multipla, ma svariati trattamenti sono disponibili per pazienti colpiti dalla malattia. Lo scopo di questi trattamenti è il rallentamento del decorso della malattia tramite l'utilizzo di farmaci antiinfiammatori e/o immunosoppressori. Spesso per i pazienti sono riscontrati importanti effetti collaterali dovuti a queste terapie. Inoltre, frequentemente, per il personale clinico la scelta del trattamento risulta difficoltosa a causa dell'imprevedibilità della patologia.

Al fine di raggiungere una miglior comprensione della malattia, fornire un supporto alla decisione clinica, migliorare la qualità di vita dei pazienti e diminuire i costi per le aziende ospedaliere, è indispensabile conoscere l'evoluzione della Sclerosi Multipla per ogni specifico paziente.

In questo lavoro di tesi, l'intento è stato quello di sviluppare diversi modelli predittivi, tramite tecniche di Machine Learning, con l'obiettivo di determinare se un paziente, a distanza di 6, 12 o 18 mesi dall'inizio dello studio, sosterrà un peggioramento della patologia o meno.

I dati forniti in input ai modelli, temporalmente datati all'inizio dello studio (giorno zero), sono stati estratti dal database MSOAC Placebo, inclusivo di informazioni riguardanti 2465 pazienti affetti da Sclerosi Multipla. Da esso, sono state ricavate 20 feature di carattere demografico e clinico.

Quattro definizioni di "peggioramento" della malattia sono state implementate per i modelli, tutte fondate su una variazione temporale dello score EDSS (Expanded Disability Status Scale).

In totale sono stati creati 72 modelli predittivi. Ciascun modello viene allenato su un Training set (80% dei dati) e successivamente validato su un Test set (20% dei dati). Per tutti i modelli è stata eseguita una normalizzazione delle variabili (tra zero ed uno) consecutiva all'imputazione dei valori mancanti (NAs) effettuata tramite il metodo MICE (Multivariate Imputation by Chained Equations). L'algoritmo con il quale sono stati implementati i modelli (nel linguaggio R) è quello di Regressione Logistica: in tal modo, è stata realizzata una classificazione binaria nelle classi "Worsening" (se un paziente vede l'aggravarsi della patologia) e "Non Worsening" (se per un paziente la malattia è stabile o migliora nel tempo di previsione). Una Stepwise Feature Selection di tipo backward e alcune tecniche di bilanciamento (Over-sampling, Both Under-sampling/Over-sampling e pesatura) sono state applicate per taluni modelli.

Le performance dei modelli sono riportate in termini del valore AUC (Area Under the Curve): 0.74 è il miglior risultato per predizioni a 6 mesi dall'inizio dello studio; 0.71 è il più alto risultato ottenuto per predizioni a 12 e 18 mesi dall'inizio dello studio. In ciascuno dei 3 modelli con le maggiori prestazioni (uno per ogni distinto arco di previsione) sono state implementate tecniche di bilanciamento. Sono stati inoltre identificati i migliori predittori per i modelli con le più alte performance: lo score EDSS e il Nine Hole Peg Test (test funzionale per la valutazione della destrezza manuale dei pazienti) sono risultate le variabili maggiormente predittive.

Le performance predittive dei modelli sviluppati sono dunque risultate significativamente migliori del classificatore random, tuttavia non sufficientemente elevate da consentire l'utilizzo nella pratica clinica dei modelli. Nonostante ciò, questa tesi ha confermato la grande potenzialità dell'applicazione di modelli predittivi per prevedere il decorso della Sclerosi Multipla. Possibili sviluppi futuri riguardano l'utilizzo di modelli non lineari e l'inclusione nei modelli di altre variabili potenzialmente predittive, ad esempio variabili estratte dalle immagini di risonanza magnetica cerebrale purtroppo non disponibili nel dataset utilizzato per questa tesi.

Capitolo 1

Sclerosi Multipla e modelli predittivi

1.1 Premessa

La Sclerosi Multipla (in inglese “Multiple Sclerosis”, abbreviata tramite l’acronimo “MS”) è una malattia neurodegenerativa, la più comune tra i giovani adulti (20-40 anni) [1]. Attualmente, si stima che circa 2,8 milioni di persone siano affette da questa patologia, con un’incidenza di 36 individui ogni 100.000. Il rischio di essere affetti dalla malattia risulta ampiamente sbilanciato tra i sessi. In media, il sesso femminile ha una probabilità doppia di contrarre la patologia rispetto al sesso maschile [2].

Nonostante sia tra le malattie più studiate in ambito neurologico, la Sclerosi Multipla risulta ancora tra le patologie più imprevedibili ed eterogenee. Negli ultimi vent’anni, numerosi trattamenti e farmaci sono stati introdotti nel mercato, al fine di rallentare il decorso della malattia, ma, ad oggi, non ne esiste ancora una cura. Inoltre, le terapie comportano spesso pesanti effetti collaterali oltre ad avere costi elevati, che in molte occasioni non si allineano agli standard necessari di “cost-effectiveness” [3]. Nel 2020, in Europa, il costo totale annuale per paziente si assestava in media a 40.300 euro. In generale, maggiore è la severità della malattia e maggiore è il costo associato [4].

Per i motivi sopra elencati, negli ultimi due decenni, molteplici studi sono stati effettuati con l’intento di sviluppare modelli predittivi in grado di prevedere il tipo di decorso che potrà avere uno specifico paziente affetto dalla malattia. Ciò è essenziale per aiutare il personale clinico a decidere quale trattamento affidare a quale paziente. In tal modo, si ha la possibilità di intervenire con le terapie in maniera preventiva,

migliorando la qualità di vita dei pazienti e diminuendo conseguentemente i costi per le aziende ospedaliere.

1.2 Sclerosi Multipla

La Sclerosi Multipla è una malattia cronica autoimmune e infiammatoria che coinvolge il sistema nervoso centrale (CNS, Central Nervous System), caratterizzata da demielinizzazione e danneggiamento transitorio o permanente degli assoni [3,5]. L'incidenza della malattia è ridotta per individui al di sotto dei 18 anni, anche se i casi accertati tra bambini sono più che quadruplicati dal 2013 al 2020 [2]. Il picco di incidenza è tra i 20 e 40 anni (media di 30 anni), con un'insorgenza della malattia in anticipo di 2-5 anni per il sesso femminile [6]. Il tasso di incidenza della Sclerosi Multipla decresce superati i 50 anni.

1.2.1 Eziologia

L'eziologia della Sclerosi Multipla è ancora ad oggi sconosciuta, ma è ormai accertato che la malattia sia causata da una combinazione di fattori ambientali e genetici [5]. Tra i fattori di rischio ambientali, attualmente 5 sono quelli confermati dalla comunità scientifica:

1. Latitudine: maggiore è la distanza dall'equatore, maggiori sono le probabilità di essere affetti dalla malattia; l'incidenza di questa correlazione è però diminuita nelle ultime decadi. La latitudine implica inoltre la relazione inversamente proporzionale tra rischio di contrarre la malattia e apporto di luce solare/vitamina D [5];
2. Migrazioni: il rischio di contrarre la malattia si modifica unicamente se la migrazione avviene in età giovanile. In tal caso, se ci si sposta da regioni ad "alto rischio" a regioni a "basso rischio", la probabilità di ammalarsi di Sclerosi Multipla si riduce. Viceversa, se la migrazione avviene nel verso opposto il rischio di malattia è maggiore [3,7];
3. Fumo: fumare aumenta il rischio di contrarre la Sclerosi Multipla. Questo vizio è stato anche associato alla maggior probabilità di conversione a fenotipi più severi della malattia e alla progressione della disabilità per i pazienti [8,9];
4. Epstein-Barr virus: è il virus responsabile della Mononucleosi infettiva, oltre che ad alcune tipologie di tumori epiteliali. Individui che sono stati infettati dal

patogeno hanno un rischio maggiore di soffrire di Sclerosi Multipla, soprattutto se l'infezione avviene dopo l'adolescenza [3,5];

5. Comorbidità vascolari: problemi di tipo vascolare comportano una progressione nella disabilità indotta dalla malattia [10].

Per quanto riguarda i fattori di rischio genetici, la malattia è stata associata alla regione HLA (Human Leukocyte Antigen), situata nel cromosoma 6 (6p21); Mutazioni nella regione comportano modifiche correlate positivamente o negativamente allo sviluppo della malattia [5].

La Sclerosi Multipla segue una ricorrenza familiare, in media, del 20%. Tra parenti di primo grado il rischio relativo è del 2.77%, tra familiari di secondo grado dell'1% e di terzo grado dello 0.88%. Nonostante ciò, la Sclerosi Multipla non viene considerata una malattia ereditaria [2,5].

1.2.2 Fisiopatologia

La fisiopatologia della Sclerosi Multipla non è ancora del tutto accertata. Si è però certi che la malattia abbia inizio a causa di un antigene che provoca la risposta immunitaria dell'organismo. Le cellule T situate nella periferia del sistema nervoso centrale oltrepassano la barriera ematoencefalica, provocando, assieme a macrofagi e cellule B, l'infiammazione. Quest'ultima, induce la creazione delle "placche" (dalle quali uno dei nomi alternativi della malattia stessa: "Sclerosi a Placche") e attiva la produzione di anticorpi da parte delle plasmacellule; il risultato finale è quindi il danneggiamento transitorio o permanente dei tessuti nervosi, in particolare della sostanza bianca [5, 11].

1.3 Sintomi

La Sclerosi Multipla si manifesta secondo una vasta gamma di sintomi, dovuti alla progressione della malattia o a causa di una ricaduta del paziente. Una ricaduta, o attacco (in inglese "relapse"), viene definita come "un episodio di disturbo neurologico tipico della Sclerosi Multipla di durata non inferiore alle 24 ore, a causa del quale sono accertate lesioni demielinizzanti e infiammatorie" [12]. Affinché due attacchi siano considerati distinti tra loro, l'inizio del primo deve distare temporalmente almeno 30 giorni dall'esordio del secondo.

I principali sintomi riscontrati in pazienti affetti dalla patologia sono:

- Disturbi di tipo cognitivo, motorio ed emisensoriale

- Depressione e pensieri suicidi
- Diplopia, oscillopsia e perdita della vista
- Tremore e vertigini
- Goffaggine e scarso equilibrio
- Deglutazione alterata
- Debolezza, fatica e dolore
- Rigidità e spasmi
- Disfunzione erettile e vescicale

1.4 Diagnosi

Ad oggi la Sclerosi Multipla viene diagnosticata tramite un insieme di criteri sintetizzati con il nome di “McDonald Criteria”. Proposti nel 2000 dall’International Panel on MS diagnosis, questi criteri si fondano sulla dimostrazione della disseminazione di lesioni, tipiche della malattia, nello spazio (DIS, Dissemination In Space) e nel tempo (DIT, Dissemination In Time). Oltre che su parametri clinici, principalmente fondati sul numero di attacchi che un paziente ha avuto, la diagnosi si basa anche su dati paraclinici. Le analisi paracliniche vengono effettuate tramite 3 principali strumenti:

- MRI: la risonanza magnetica consente di identificare lesioni nel cervello e nel midollo spinale. I criteri seguiti per l’interpretazione dei dati ricavati con questo strumento sono quelli di Barkhof et al. [13] e Tintorè et al. [14]. Questa tecnica di imaging, nella maggior parte dei casi, possiede alte sensibilità e specificità;
- CSF: L’analisi del liquido cerebrospinale (CerebroSpinal Fluid) riporta informazioni sulla natura infiammatoria e immunitaria delle lesioni. E’ un fondamentale strumento di supporto, soprattutto nei casi in cui la MRI ha bassa specificità (per esempio per un paziente anziano, in cui le lesioni possono essere ricondotte a comorbidità). Anomalie del liquido cerebrospinale vengono identificate tramite la presenza di bande oligoclonali IGg (Immunoglobuline G) o dalla presenza di un elevato indice di IGg [12];

- VEP: il Visual Evoked Potential, è lo studio di potenziali evocati visivi, strumento di supporto per la valutazione di lesioni al sistema visivo umano.

Grazie ai McDonald Criteria del 2000, la diagnosi della patologia è stata semplificata per il personale clinico, con ancora però alcune limitazioni: la robustezza del metodo era riscontrata principalmente se applicato a ristrette etnie (in particolare “bianco caucasico”) ed età (tra i 10 e i 59 anni). Per oltrepassare questi ostacoli, sono state effettuate tre distinte revisioni del criterio di McDonald: la revisione del 2005 da parte di Polman et al. [15], che mantiene le fondamenta del criterio originale, la revisione del 2010, sempre da parte di Polman et al. [16] ed infine l’ultima revisione, datata al 2017, da parte di Alan J Thompson et al. [17]. Quest’ultima è la versione dei criteri diagnostici attualmente seguiti dalla maggior parte della comunità clinica e scientifica. I nuovi criteri hanno portato ad una semplificazione della diagnosi, che diventa più pratica e veloce, senza una riduzione della specificità, della sensibilità e dell’accuratezza totale. Inoltre, rispetto alla prima versione proposta, questi nuovi criteri possono essere applicati per la diagnosi di pazienti asiatici e latino americani, oltre che per soggetti pediatrici [16,17].

1.5 Fenotipi

I principali fenotipi della Sclerosi Multipla possono essere suddivisi in 4 macrocategorie: Clinically Isolated Syndrome (CIS), Relapse-Relmitting MS (RRMS), Primary Progressive MS (PPMS) e Secondary Progressive MS (SPMS) [18,19]. Prima di entrare nel dettaglio di ciascuna forma della malattia, è bene chiarire la terminologia che in seguito viene utilizzata. Viene considerato *peggioramento* (“Worsening”) della malattia laddove è documentato un incremento delle disabilità e/o disfunzioni neurologiche del paziente. Con il termine malattia *attiva* si intende una attività della patologia di durata non inferiore a un anno, confermata da riscontri clinici e tramite MRI. Il termine *progressione* della malattia viene preferibilmente utilizzato per le varianti progressive (PPMS e SPMS) e viene interpretato come un avanzamento della patologia attestato da dati clinici e paraclinici in un minimo arco temporale di un anno e che comporta un peggioramento irreversibile delle funzioni neurologiche [19,20].

In generale, l’evoluzione clinica della malattia può essere ricondotta per la maggior parte dei pazienti a 3 fasi [20]:

1. Fase ad alto rischio;
2. Fase Relapse-Relmitting;

3. Fase Progressiva.

1.5.1 Clinically Isolated Syndrome (CIS)

Il fenotipo CIS è caratterizzato da un primo e unico evento neurologico causato da demielinizzazione in concomitanza ad infiammazione, per il quale però non sono ancora stati rispettati i requisiti di “disseminazione nel tempo” delle lesioni al CNS [18,19,20,21]. Questa tipologia di Sclerosi Multipla può manifestarsi sia localmente nel CNS (monofocale), sia in maniera diffusa (multifocale) [21]. La forma CIS può essere suddivisa in due sottogruppi:

- Solitary Sclerosis (SS): con questo termine si fa riferimento a una forma che colpisce pazienti che non hanno rispettato i requisiti MRI per la diagnosi della malattia [13,14], ma hanno lesioni (tipiche della Sclerosi Multipla) al CNS non riconducibili ad altre malattie [20];
- Single Attack MS (SAMS): con questa espressione si individua un paziente che ha avuto un singolo attacco neurologico che tramite MRI rispetta i requisiti prestabiliti dai McDonald criteria, ovvero può essere immediatamente considerato una relapse [20].

1.5.2 Relapse-Relmitting Multiple Sclerosis (RRMS)

La forma Relapse-Relmitting (in italiano Recidivante-Remittente), con la quale la maggior parte dei pazienti viene diagnosticata, si manifesta tramite diversi episodi di attacchi in differenti parti del CNS e separati nel tempo, rispettando così le fondamenta dei criteri diagnostici di McDonald [19,20,21]. Questi episodi sono poi seguiti da periodi di completa o parziale remissione delle funzioni neurologiche. Anche questo fenotipo della malattia, come il CIS, può essere suddiviso, in un arco temporale solitamente di un anno, in: RRMS attiva, se sono presenti ricadute, o RRMS non attiva, se il paziente gode di un periodo “silente” della patologia [19,21]. Inoltre, una ulteriore suddivisione scinde il fenotipo Recidivante-Remittente in: RRMS con peggioramento, ovvero se il periodo di attività è seguito da un parziale, e non totale, recupero delle capacità neurologiche; oppure RRMS senza peggioramento, quando il recupero successivo agli attacchi è completo [21].

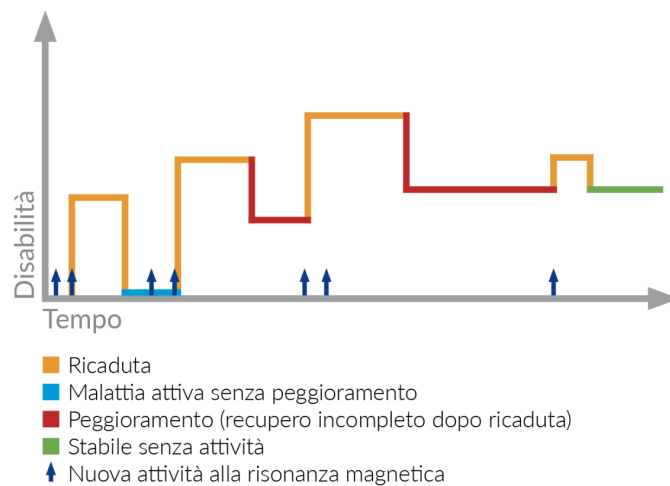


Figura 1.1: Andamento della Sclerosi Multipla nel caso del fenotipo RRMS. Sorgente: aism.com

1.5.3 Secondary Progressive Multiple Sclerosis (SPMS)

La forma Secondariamente Progressiva della Sclerosi Multipla si manifesta dopo un tempo mediano di 19 anni per all'incirca il 75% dei pazienti affetti dal fenotipo RRMS [5]. La SPMS è perciò considerata l'evoluzione del fenotipo Recidivante-Remittente, sebbene non ci sia un passaggio netto nel tempo tra le due forme: la conversione risulta graduale [18,20]. La diagnosi della forma SP viene effettuata in maniera retrospettiva, visionando il progressivo peggioramento della malattia per il paziente. Questa tipologia di Sclerosi Multipla è caratterizzata da un peggioramento delle disabilità continuo nel tempo, assieme al quale si verificano sporadici attacchi nel 40% dei pazienti, soprattutto nella fase iniziale [5]. La SPMS può essere suddivisa in quattro categorie [18,19]:

- SPMS attiva e con progressione: il paziente ha ricadute in contemporanea alla progressione della malattia;
- SPMS attiva ma senza progressione: il soggetto ha avuto una ricaduta ma senza peggioramento della disabilità;
- SPMS non attiva ma con progressione: il paziente non ha attacchi ma è rilevante il peggioramento della malattia (es. le difficoltà di camminata sono aumentate);
- SPMS non attiva e senza progressione: la patologia è stabile.

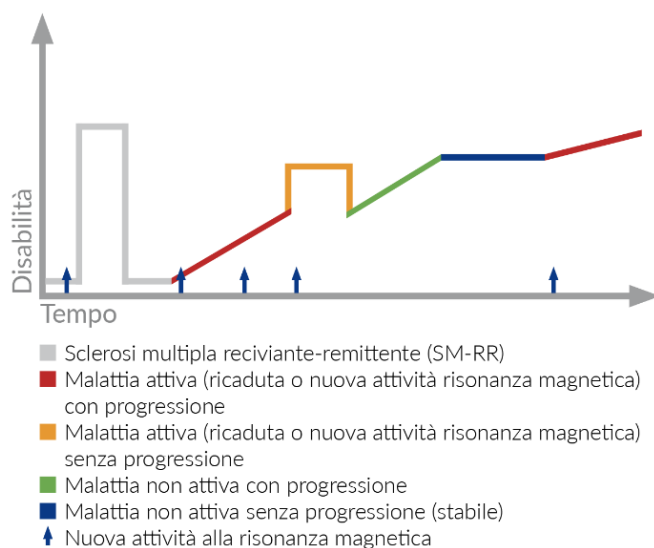


Figura 1.2: Andamento della Sclerosi Multipla nel caso del fenotipo SPMS. Sorgente: aism.com

1.5.4 Primary Progressive Multiple Sclerosis (PPMS)

Il fenotipo Primariamente Progressivo di Sclerosi Multipla si manifesta nella forma progressiva (simile a quella del fenotipo SPMS) sin dall'esordio della malattia [18,19]. Questa forma della patologia colpisce il 15% dei pazienti, con una età media di esordio della malattia che si attesta attorno ai 40 anni [5]. La progressione della disabilità che la patologia comporta è più severa rispetto al caso di pazienti nei quali vi è stata la conversione del fenotipo RRMS a SPMS: per pazienti PPMS con importanti difficoltà motorie e psichiche, il sostegno per compiere 100 m a piedi è necessario dopo 7 anni; per pazienti SPMS (con identiche difficoltà motorie e cognitive) il supporto per la camminata è impiegato dopo 12.5 anni [5]. Come per il fenotipo SMPS, anche per la forma Primariamente Progressiva esistono 4 differenti tipologie. Queste sono: attiva e progressiva, attiva ma non progressiva, non attiva ma progressiva e non attiva e non progressiva [18,19]. I significati di questi sottotipi corrispondono a quelli del fenotipo Secondariamente Progressivo elencati nel sottocapitolo 1.5.3.

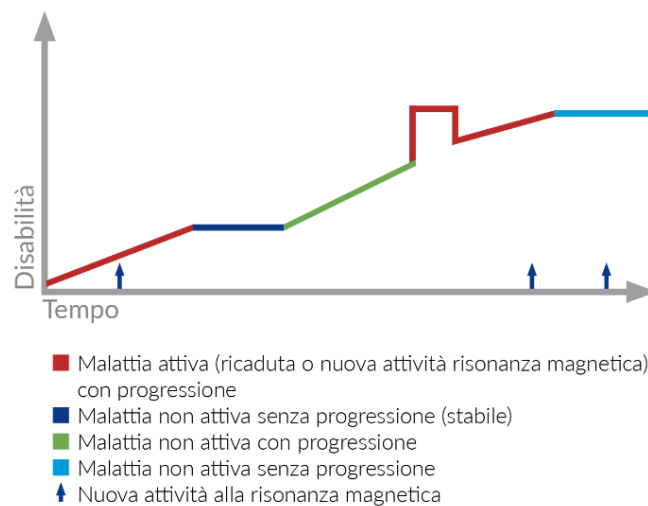


Figura 1.3: Andamento della Sclerosi Multipla nel caso del fenotipo PPMS. Sorgente: aism.com

1.6 Trattamenti

Come anticipato nella sezione 1.1, ad oggi non esiste ancora una cura per la Sclerosi Multipla. Nonostante ciò, numerose terapie sono attualmente somministrate ai pazienti con l'intento di: diminuire la severità degli attacchi, ridurre la frequenza, e rallentare il decorso della malattia, ovvero attenuarne il peggioramento e, conseguentemente, alleviare i sintomi [22,23]. La maggior parte di questi trattamenti sono rivolti a pazienti affetti dal fenotipo RRMS; in alcuni casi, sono adoperati per pazienti colpiti dalle forme CIS e SPMS. Nessuna terapia è attualmente disponibile per trattare il fenotipo PPMS [5,22]. I farmaci impiegati al giorno d'oggi possono essere raggruppati in 2 macrocategorie:

- Farmaci di prima linea o terapie di base;
- Farmaci di seconda linea o terapie di escalation;

Nella prima categoria sono inclusi farmaci a basso rischio, con leggeri effetti collaterali nella maggior parte dei casi. Nella seconda, sono presenti farmaci più potenti, con effetti collaterali potenzialmente più rischiosi [5].

1.6.1 Terapie di base

- *Interferoni Beta*: gli Interferoni Beta sono le sostanze più utilizzate per la terapia di prima linea; essi sono conosciuti e applicati da più di due decenni a pazienti affetti dal fenotipo RRMS [22]. Impiegati in due forme differenti,

Interferoni Beta-1a e Interferoni Beta-1b, gli Interferoni Beta sono dei polipeptidi con proprietà antiinfiammatorie; hanno inoltre effetti inibitivi sulla proliferazione e migrazione dei linfociti T [24]. Per pazienti diagnosticati con forma Recidivante-Remittente, gli Interferoni Beta hanno mostrato una riduzione dell'ARR (Annualized Relapse Rate, ovvero il numero di ricadute che un soggetto sperimenta in un arco temporale di un anno) attorno al 30%-34%, diminuendo anche il peggioramento delle disabilità per il soggetto. Studi hanno riportato un rischio ridotto, grazie agli Interferoni Beta, di nuove attività per pazienti con fenotipo CIS. Anche individui affetti dal fenotipo SPMS attivo risentono di benefici dall'uso di questo trattamento [24]. Possibili effetti collaterali, dovuti all'utilizzo di queste sostanze, sono: dolori muscolari, febbre, mal di testa, mal di schiena, reazioni nel sito di iniezione, insufficienza epatica e produzione di anticorpi neutralizzanti [21,24].

- *Glatiramer Acetato*: in ordine di frequenza di utilizzo, si trova poi il Glatiramer Acetato. Questa sostanza, costituita da peptidi sintetici, ha effetti antiinfiammatori (conduce a un rilascio di citochine antiinfiammatorie). In clinica, il Glatiramer Acetato ha dimostrato una riduzione dell'ARR di circa il 29% per pazienti colpiti dal fenotipo RRMS. In pazienti affetti da CIS, il farmaco è in grado di prolungare il tempo di comparsa del secondo attacco. Questa terapia non viene somministrata invece a pazienti che convivono con i fenotipi progressivi della malattia [23,24]. Nel 65% dei casi i pazienti sperimentano reazioni nel sito di iniezione. L'effetto collaterale più rilevante, poiché permanente, è la lipoatrofia [24].
- *teriflunomide*: il teriflunomide, terzo farmaco per frequenza di utilizzo nelle terapie di base, è un agente immunomodulatore il cui obiettivo ultimo è la riduzione del numero e della capacità proliferativa dei linfociti. Il farmaco, per pazienti RRMS, ha dimostrato un calo dell'ARR del 31%-36% [24]. Questo agente ha portato a miglioramenti delle condizioni per pazienti CIS in termini di tempo aumentato tra la prima e la seconda relapse, ma non viene utilizzato come rimedio delle forme progressive.
- *dimetilfuramato*: in ambito clinico si possono trattare pazienti affetti dalla forma RRMS con l'utilizzo del dimetilfuramato, un agente immunomodulatore con proprietà antiinfiammatorie. In media, questa sostanza comporta una diminuzione dell'ARR del 44%-53% [24]. L'impiego di tale terapia richiede tuttavia di prestare estrema attenzione ai possibili effetti collaterali, il peggiore tra i quali è la leucoencefalopatia multifocale progressiva [21].

1.6.2 Terapie di escalation

Le terapie di seconda linea, più potenti ma con maggiori e peggiori effetti collaterali, si fondano principalmente sui seguenti principi attivi:

- *Fingolimod*: immunomodulatore e immunosoppressore, questa sostanza inibisce la capacità di passaggio di linfociti B e T dai linfonodi al CNS [21, 22]. In ambito clinico ha dimostrato una riduzione dell'ARR pari al 48%-55% [24]. Attualmente, il farmaco che include questo principio attivo, Gilenya, viene affidato unicamente a pazienti con forma RRMS. Effetti collaterali importanti spaziano da asistolia e bradicardia sostenuta, fino alla morte dovuta ad aritmia ventricolare [5];
- *Natalizumab*: è un anticorpo monoclonale che previene l'adesione di cellule T e B all'endotelio infiammato. Inoltre, ne impedisce la migrazione nel CNS [24]. La sua efficacia nel diminuire il numero annuale di relapses è notevole: si registra un ARR diminuito, in media, del 68% per pazienti affetti da forma RRMS della malattia (unici a poter godere di questa terapia al giorno d'oggi). Il rischio maggiore che questo farmaco comporta è lo sviluppo di leucoencefalopatia multifocale progressiva [21, 24];
- *Mitoxantrone*: derivato sintetico dell'antacenedrione, il Mitoxantrone interagisce con il DNA nucleare sopprimendo la proliferazione di diverse cellule immunitarie, tra le quali cellule B, cellule T e macrofagi [21,24]. La sua efficacia in termini di riduzione dell'ARR è elevata: si attesta una diminuzione del 66% sia per pazienti affetti da fenotipo RRMS, sia per individui con forma SPMS [5,24]. Effetti collaterali gravi includono leucemia mieloide acuta, sindrome mielodisplastica e leucemia acuta [21].

1.7 Modelli predittivi per la Sclerosi Multipla

Gli algoritmi di Machine Learning (ML) sono in grado di creare modelli predittivi che necessitano di un minimo intervento umano. Questi modelli hanno la capacità di prevedere, utilizzando dati del presente, un prestabilito output in un arco temporale futuro. Talvolta, quando si lavora in ambito medico, i modelli sviluppati, vengono nominati “prognostici”: lo scopo di questi modelli diviene per cui la previsione del decorso o dell’esito di una patologia. L’abilità di predire il peggioramento della MS, condurrebbe a molteplici vantaggi, tra i quali:

1. una migliore comprensione della patogenesi della malattia [25];
2. agevolare lo sviluppo di trattamenti, sia nuovi che preesistenti [25];
3. da un lato, evitare di conferire trattamenti gravosi a pazienti affetti da forme “lievi” della malattia; dall’altro, scongiurare terapie moderate per soggetti predisposti a forme più “aggressive” della patologia [26];
4. ottimizzare, in generale, la gestione e il monitoraggio dei pazienti [27, 28];
5. anticipare i trattamenti per pazienti che necessitano di terapie preventive, migliorando di conseguenza il rapporto rischi/benefici e la qualità di vita dei soggetti [25, 27];
6. ridurre il tempo richiesto per svolgere i trial clinici e diminuire il numero di pazienti sottoposti a terapie sperimentali [25];
7. informare i pazienti riguardo al loro futuro e alle decisioni da intraprendere [25];
8. abbassare i costi della sanità, conseguenza dei punti precedenti.

Lo stato dell’arte, in quest’ambito, include svariati studi di applicazione di tecniche di ML con l’intento di realizzare previsioni del peggioramento/progressione nel tempo della Sclerosi Multipla. Le metodologie adottate per lo scopo possono essere categorizzate in:

- metodi di regressione: hanno come obiettivo ultimo quello di predire, per ciascun individuo, il valore esatto di una variabile continua in un punto temporale futuro rispetto ai dati utilizzati raccolti in uno specifico momento antecedente;

- metodi di classificazione: perseguono lo scopo di classificare i pazienti in distinte classi, solitamente due (classificazione binaria, per esempio pazienti che peggioreranno vs pazienti che saranno stabili in futuro);
- analisi di sopravvivenza: studiano il tempo intercorso tra un evento iniziale (solitamente l'inizio dello studio) e un avvenimento finale (tradizionalmente la morte dell'individuo).

Nell'applicazione di metodologie di ML nell'ambito dell'MS, la maggior parte dei modelli si fonda come precedentemente detto sulla ricerca del peggioramento o progressione della patologia. Numerose sono le scelte attuabili per dimostrare l'aggravarsi della Sclerosi Multipla, da cui il numero elevato di outcomes definiti per i modelli. Alcuni autori, per esempio, tentano di predire nel futuro il valore esatto di un test funzionale svolto dai pazienti; altri ambiscono alla previsione dell'aumento della fatica negli individui [29]. In letteratura, due sono gli outcomes maggiormente scelti per cogliere l'avanzamento della malattia:

- conversione o tempo impiegato per la conversione da fenotipo RRMS a fenotipo SPMS [30,31,32,33];
- peggioramento della malattia basato sulla variazione del valore di score che misurano le capacità funzionali del paziente (come per esempio la Expanded Disability Status Scale, una scala di valutazione delle disabilità in diversi sistemi funzionali del paziente - vedi Sezione 1.8)

In questa tesi, l'attenzione è posta sulle metodologie di Classificazione in combinazione alla ricerca dell'outcome "variazione del valore EDSS nel tempo".

1.8 EDSS e metodi di classificazione

L'EDSS (Expanded Disability Status Scale) è un sistema di scoring introdotto nel 1983 dal professore di neurologia John F. Kurtzke come estensione della scala DSS (Disability Status Scale) precedentemente utilizzata [34]. Lo scopo della scala EDSS è la valutazione della disabilità di pazienti affetti da Sclerosi Multipla. L'intervallo di variazione dei valori è contenuto in un intervallo tra 0 e 10.

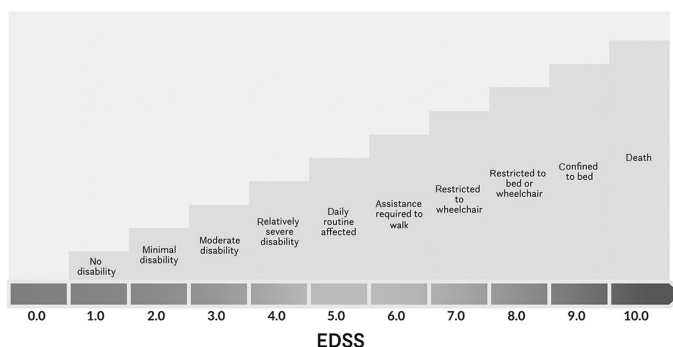


Figura 1.4: Expanded Disability Status Scale (EDSS). Sorgente: cambridge.org

Un paziente con valore 0 ha un esame neurologico normale. Man mano che la disabilità progredisce, il valore affidato al soggetto aumenta (es. EDSS pari a 6 identifica un individuo che cammina con il sostegno di un bastone) fino al valore 10, corrispondente alla morte dell'individuo dovuta alla malattia. Questo score è calcolato in base al risultato di 8 sistemi funzionali (FS, Functional Systems): Pyramidal Functions, Cerebellar Functions, Brain Stem Functions, Sensory Functions, Bowel and Bladder Functions, Visual (or Optic) Functions, Cerebral (or Mental) Functions e Other Functions. Ognuno di questi sistemi funzionali è indipendente dagli altri ma assieme esprimono tutte le alterazioni neurologiche indotte dalla patologia [35].

Oltre ad essere utilizzato come routine nella pratica clinica per tenere traccia dello stato di disabilità del paziente, lo score EDSS è un parametro utilizzato in quasi tutti gli studi di ML applicato alla Sclerosi Multipla. Spesso la sua presenza è fondamentale per garantire buone prestazioni dei modelli. In alcuni lavori, tuttavia, il suo uso è criticato, venendo considerato un parametro superficiale, non sufficientemente accurato per il calcolo della disabilità del paziente [5,25,28]. Nonostante ciò, come detto nella sezione 1.7, una variazione del valore dello score nel tempo viene abitualmente utilizzata per definire un peggioramento della malattia. Il criterio per stabilire quanto debba essere la variazione del valore EDSS per determinare il peggioramento, non è univoco; non esiste il così detto “gold standard”. Ogni studio definisce il proprio “peggioremento basato sul Δ EDSS” (ovvero sulla differenza tra

valore EDSS in un momento futuro e valore EDSS all’inizio dello studio), il quale si discosta di un certo grado dalla definizione introdotta da Kurtzke, che definisce che un aumento di 0.5 sulla scala EDSS indica che il paziente è peggiorato [35]. Definito il “peggioramento” del paziente, si procede, in ambito di classificazione, a separare i soggetti in esame in due classi: “Worsening” (ovvero “peggioramento”), pazienti per cui la malattia è progredita in uno specifico arco temporale (ovvero pazienti che hanno un aumento del valore EDSS, quindi un Δ EDSS positivo o superiore ad una certa soglia fissata), o “Non Worsening” (ovvero “non peggioramento”), pazienti per cui nello stesso periodo non sono stati riscontrati peggioramenti della patologia (ovvero soggetti in cui non si riscontra un aumento del valore EDSS, per i quali si ha un Δ EDSS minore o uguale a zero).

Nelle sezioni seguenti vengono presentati alcuni tra i più rilevanti articoli che si propongono di classificare i pazienti nelle due classi, utilizzando come output di modello, e quindi separatore tra classi, il Δ EDSS nel tempo.

1.8.1 Modello secondo Zhao et al., 2017

Lo studio di Zhao et al. del 2017 [36] fonda la sua analisi sulla classificazione di pazienti nelle classi “Worsening”, se un soggetto ha un aumento in 5 anni di almeno 1.5 sulla scala EDSS e “Non Worsening”, per i rimanenti individui. Utilizzando il database CLIMB, inclusivo di 1693 pazienti, i modelli classificatori vengono costruiti tramite due differenti algoritmi: SVM (Support Vector Machines), con funzione di Kernel sia lineare che non, e LR (Logistic Regression). Le feature, ovvero le variabili a disposizione tra i dati di CLIMB adoperate come input per i modelli, sono demografiche, cliniche e paracliniche. I modelli realizzati classificano i pazienti sulla base di dati registrati alla baseline (fissata in corrispondenza dell’inizio dello studio), baseline con aggiunta di dati a 12 mesi, baseline con aggiunta di dati a 12 e 24 mesi. I risultati trovati utilizzando unicamente i dati alla baseline riportano un’accuratezza totale media del 63%. Per ottenere risultati di accuratezza totale superiore all’80% è stata necessaria l’aggiunta di dati a 12 e 24 mesi dall’inizio dello studio.

1.8.2 Modello secondo Zhao et al. 2020

Lo studio di Zhao et al. del 2020 [28], come nello studio [36], suddivide i pazienti in base all’aumento ≥ 1.5 o meno del valore EDSS in 5 anni. Anche qui le feature utilizzate sono demografiche, cliniche e paracliniche. Oltre agli algoritmi SVM e LR, già applicati nel 2017 [36], in questo studio Zhao et al. sviluppano i modelli

utilizzando l’algoritmo RF (Random Forest), oltre che adoperando un approccio di ensemble learning, ovvero utilizzando più algoritmi in un unico modello. In questo secondo caso, gli algoritmi utilizzati sono: XGboost, Gradient Boosting Machine (GBM), and Meta-learner (unione di SVM, LR e RF). In questo studio il database è costituito da 724 soggetti, con ulteriori 400 utilizzati per la validazione esterna dei modelli. In generale, sono risultati migliori gli algoritmi “ensemble”, rispetto ai classici. Il miglior risultato è stato ottenuto tramite l’algoritmo Meta-learner, raggiungendo un valore AU-ROC (Area Under the Receiver Operating Characteristic Curve) pari a 0.83 sul dataset di validazione.

1.8.3 Modello secondo Rocca et al., 2021

Nello studio di Rocca et al. del 2021 [27], la classificazione è determinata in base al valore EDSS che ciascun paziente ha all’inizio del trial clinico, oltre che alla sua variazione nel tempo. In particolare:

- pazienti con valore EDSS iniziale pari a 0 sono assegnati alla classe “Worsening” se e solo se la loro variazione di EDSS dopo 6,4 anni è almeno di 1.5;
- pazienti con valore EDSS alla baseline compreso tra 0.5 e 6 sono assegnati alla classe “Worsening” se e solo se hanno una variazione di EDSS a distanza di 6.4 anni di almeno 1 punto;
- pazienti “critici”, ovvero con EDSS alla baseline maggiore di 6 sono assegnati alla classe “Worsening” se e solo se la variazione di EDSS a 6.4 anni dalla baseline è superiore o uguale a 0.5.

Utilizzando dati demografici, clinici e paraclinici di 233 pazienti, i modelli predittivi sono stati sviluppati con l’algoritmo DT (Decision Tree). Dopo aver testato diverse combinazioni di variabili, le performance ottimali sono state individuate con l’utilizzo della totalità dei parametri. Il risultato in termini AUC in tal caso è risultato pari a 0.76.

1.8.4 Modello secondo Pinto et al., 2020

Il modelli sviluppati da Pinto et al. nel 2020 [37] considerano la presenza di un peggioramento della malattia nel momento in cui la differenza tra il valore calcolato come media del valore EDSS tra il sesto e il decimo anno dalla baseline e il valore EDSS all’inizio dello studio è maggiore o uguale a 3. Disponendo di dati di 212

pazienti, gli autori hanno utilizzato variabili demografiche, cliniche e paracliniche per costruire 4 diversi classificatori basati sugli algoritmi SVM, k-NN (k-Nearest Neighbour), DT e Linear Regression. I migliori risultati nel predire il peggioramento della malattia corrispondono a valori di AUC pari a 0.89 per le predizioni a 6 anni dalla baseline e AUC pari a 0.85 per le predizioni a 10 anni dalla baseline.

1.8.5 Modello secondo Law et al., 2019

Law et al. nel 2019 [25] hanno sviluppato modelli di classificazione fondati sulla seguente definizione di peggioramento della malattia:

- un paziente con valore EDSS alla baseline ≤ 5.5 viene assegnato alla classe “Worsening” se ha un incremento a distanza di 6 mesi di almeno 1 punto sulla scala EDSS;
- un paziente con valore EDSS alla baseline ≥ 6 viene assegnato alla classe “Worsening” se ha un aumento a distanza di 6 mesi di almeno 0.5 sulla scala EDSS.

Tramite feature demografiche, cliniche e paracliniche di 485 soggetti, gli autori hanno allenato modelli basati sugli algoritmi LR, SVM lineare, DT, ensemble LR, ensemble linear SVM, RF e AdaBoost. Le performance migliori, in termini di AUC, sono state ricavate tramite l’algoritmo DT, raggiungendo un valore massimo di 0.618. Gli autori ipotizzano che la ridotta capacità predittiva dei modelli sviluppati sia dovuta alla mancanza di dati a disposizione (sia come numero di pazienti sia come quantità di variabili).

1.8.6 Modello secondo Tommasin et al., 2021

Tommasin et al. nel 2021 [38] fondano la propria analisi su una classificazione che rispetta i seguenti criteri:

- Un paziente “non critico”, con EDSS alla baseline pari a 0, viene assegnato alla classe “Worsening” se ha un incremento a 2 anni del parametro EDSS di almeno 1.5;
- Un paziente “intermedio”, con EDSS alla baseline compreso tra 1 e 5, viene assegnato alla classe “Worsening” se ha un aumento a 2 anni di almeno 1 punto sulla scala EDSS;

- Un paziente “critico”, ovvero con valore EDSS alla baseline superiore o uguale a 5.5, viene assegnato alla classe “Worsening” se ha un incremento maggiore o uguale a 0.5 nella scala EDSS a distanza di 2 anni.

Gli autori hanno impiegato dati di 163 pazienti, per i quali disponevano di feature - come nei precedenti studi - sia demografiche e cliniche sia paracliniche (MRI data). Implementando un algoritmo di RF per 1000 volte hanno ottenuto mille classificatori differenti. Il miglior classificatore ha riportato un risultato di accuratezza totale pari a 0.79.

1.8.7 Modello secondo Yperman et al., 2020

Nel 2020, Yperman et al. [39] hanno suddiviso i pazienti in “Worsening” e “Non Worsening” secondo le seguenti due definizioni di peggioramento:

- Un soggetto che ha EDSS al tempo T0 inferiore o uguale a 5.5, è assegnato alla classe “Worsening” se al tempo T1 ha avuto un aumento sulla scala EDSS di almeno 1 punto;
- Un individuo che ha EDSS al tempo T0 superiore a 5.5 viene assegnato alla classe “Worsening” se al tempo T1 ha subito una progressione di almeno 0.5 sulla scala EDSS.

L’inizio dello studio per i pazienti viene identificato come tempo “T0”. Due anni dopo l’inizio dello studio è definito come tempo “T1”. I modelli sviluppati da Yperman et al., fondati sugli algoritmi di RF e LR, utilizzano come feature di input dati di MEPs (Motor Evoked Potentials), ricavati da 642 pazienti. La miglior performance, ottenuta implementando modelli RF, raggiunge un valore AUC di 0.75.

1.9 Scopi e organizzazione della tesi

Dai modelli descritti nella sezione 1.8 emerge la mancanza di una definizione standardizzata di “peggioramento” della Sclerosi Multipla. Assieme a questo problema, si osserva spesso una cardinalità limitata di pazienti nei database utilizzati negli studi. Questa è una complicazione rilevante, poiché gli algoritmi di ML necessitano di una grande mole di dati affinché possano raggiungere buone prestazioni in termini di accuratezza e generalizzabilità delle predizioni ottenute. Per andare incontro a quest’ultima criticità, nella presente tesi, è stato richiesto l’accesso al database MSOAC Placebo, inclusivo di 2465 pazienti affetti da Sclerosi Multipla. Data invece la non unicità della definizione di “peggioramento” della malattia basata sulla variazione del parametro EDSS, in questo lavoro sono stati implementati modelli che classificano i pazienti secondo le 3 soglie più comunemente applicate al valore ΔEDSS : quella introdotta da Zhao et al. [28,36], da Rocca et al. [27] e da Law et al. [25]. Oltre a queste tre soglie, è stata proposta infine una soglia “Zero”: ΔEDSS maggiore di 0 indica che il paziente peggiora nel tempo, ΔEDSS minore di 0 indica che il paziente è stabile o migliora nel tempo. Quest’ultima soglia, nonostante sia poco rilevante dal punto di vista clinico, è stata introdotta per esplorare l’effetto del possibile sbilanciamento tra le classi.

Partendo dai dati demografici e clinici del database MSOAC Placebo, lo scopo della tesi è improntato alla creazione di molteplici modelli ai quali sono associate diverse definizioni di “peggioramento della MS”, quindi differenti tipologie di classificazione. Dopo questo primo capitolo di introduzione, descrittivo della Sclerosi Multipla e dello stato dell’arte di sviluppo di modelli in Machine Learning applicabili a pazienti affetti da Sclerosi Multipla, seguirà il Capitolo 2: dati utilizzati e pre-processing. In questo capitolo verrà descritto il database MSOAC Placebo e le operazioni di analisi eseguite per arrivare ad un dataframe contenente i dati dei soggetti alla baseline poi utilizzati come input per i modelli prognostici. Seguirà poi il Capitolo 3, inclusivo della pipeline ideata per la creazione dei modelli. In questo capitolo verrà esposta la scelta dell’orizzonte temporale di previsione, le definizioni di “peggioramento della malattia” utilizzate ed altre tecniche statistiche applicate. Per concludere verrà infine il Capitolo 4, in cui sono riportati i risultati ricavati in termini del valore AUC, oltre che una lista dei migliori predittori per i diversi modelli. Un commento sui risultati ottenuti e su come si pongono rispetto allo stato dell’arte verrà riportato nelle conclusioni della tesi (Capitolo 5).

Capitolo 2

Il database MSOAC Placebo: analisi e pre-processing

2.1 Il database MSOAC Placebo

Per sviluppare modelli predittivi del peggioramento della MS, in questa tesi sono stati utilizzati dati ricavati dal database MSOAC Placebo. Il “Multiple Sclerosis Outcome Assessments Consortium” (MSOAC) è un consorzio formatosi nel 2012, organizzato e gestito dal Clinical Path Institute (C-path) con il supporto della National Multiple Sclerosis Society (NMSS) [40]. Lo scopo del progetto MSOAC è quello di accelerare lo sviluppo di nuove terapie per la Sclerosi Multipla mediante la ricerca di strumenti innovativi per la misurazione degli outcomes nei trial clinici della MS [41]. Per generare questi strumenti, molteplici altre risorse sono state prodotte, tra le quali uno Standard per la raccolta e gestione dei dati relativi a pazienti affetti dalla malattia, creato con la collaborazione del Clinical Data Interchange Standards Consortium (CDISC). Questa grande potenzialità è stata poi impiegata per mappare i dati di 16 differenti studi clinici, fondendoli in un unico database: il MSOAC database. In questa raccolta, sono presenti dati di 14.370 pazienti, appartenenti a rami sia di trattamento che di controllo dei trial clinici [41,42]. A ricercatori qualificati è garantito l’accesso ai dati del ramo “placebo” (<https://c-path.org/programs/msoac/>) che coinvolge 2465 pazienti provenienti da 9 dei 16 trial clinici. A causa della procedura svolta per rendere i dati anonimi, non è possibile conoscere quali siano gli studi da cui provengono i pazienti del ramo placebo [43]. Come anticipato, la presente tesi utilizza appunto i dati del database MSOAC Placebo. Richiesti sul portale del C-Path Institute, vengono forniti tramite:

- Un insieme di file “.csv” contenenti i dati dei pazienti;

- Un file Word intitolato “Description of how timing is recorded in the MSOAC data”, utile per capire come viene gestita la variabile “tempo” nel database;
- Un file Excel che include uno schema riassuntivo del database.

Il database MSOAC placebo contiene unicamente dati demografici e clinici, mentre non ci sono informazioni paracliniche, per esempio MRI data. La banca dati è suddivisa in un totale di 21 domini, di cui 9 costituiti da materiale di supporto. Questi domini devono rispettare i requisiti di: contenere dati riguardanti la Sclerosi Multipla, essere costituiti da grandezze oggettive e facilmente misurabili, e comprendere dati che riflettono l’abilità nel completare funzioni utili alla vita quotidiana da parte di pazienti affetti dalla patologia [39]. Ciascun dominio è rivolto a una tipologia specifica di dati, la cui aggregazione rende il database una raccolta varia ed eterogena di informazioni che descrive sotto diversi punti di vista una classica popolazione di pazienti affetti dalla MS. Sono inclusi dati puramente demografici (es. età), dati riguardanti il fenotipo della malattia (vedi sezione 1.5), risultati di test funzionali (riscontro clinico) e questionari (riscontro dei pazienti), e misure inerenti la storia medica e le relapses che hanno avuto i soggetti [43]. Un riassunto dei domini costituenti il database MSOAC placebo è riportato nella tabella 2.1.

Tabella 2.1: Descrizione e sintesi dei domini costituenti il database MSOAC Placebo

DOMINIO	NOME COMPLETO	DESCRIZIONE
CE	Clinical Events	Dati su Relapses, sia confermate che non
CM	Concomitant Medications	Medicinali e terapie seguite prima e durante lo studio
DM	Demographics	Dati demografici dei pazienti
DS	Disposition	Consenso informato, randomizzazione e uscita anticipata dallo studio
FA	Findings About	Relapses nei tre anni precedenti allo studio o dalla diagnosi della malattia
FT	Functional tests	Test funzionali di mobilità o cognizione
MH	Medical History	Condizioni mediche prima dello studio
OE	Ophthalmic Examinations	Test sul sistema visivo (acuità)
QS	Questionnaires	Questionari e scores disabilità
RP	Reproductive System Findings	Risultati test di gravidanza
SC	Subject Characteristics	Mano dominante (destra o mancina)
SM	Subject Disease Milestones	Dati su Relapses confermate
SUPPCE	Supplemental Clinical Events	Materiale di supporto per il dominio CE
SUPPCM	Supplemental Concomitant Medications	Materiale di supporto per il dominio CM
SUPPDM	Supplemental Demographics	Materiale di supporto per il dominio DM
SUPPFA	Supplemental Findings About	Materiale di supporto per il dominio FA
SUPPFT	Supplemental Functional Test	Materiale di supporto per il dominio FT
SUPPMH	Supplemental Medical History	Materiale di supporto per il dominio MH
SUPPQS	Supplemental Questionnaires	Materiale di supporto per il dominio QS
SUPPSC	Supplemental Subject Characteristics	Materiale di supporto per il dominio SC
SUPPSM	Supplemental Subjects Disease Milestones	Materiale di supporto per il dominio SM

2.2 Analisi e pre-processing

Come anticipato nella sezione 1.9, lo scopo di questa tesi è quello di implementare dei modelli computazionali predittivi che ricevano in input dati alla baseline dei soggetti (presente) e prevedano in output la variazione del parametro EDSS nel tempo (futuro), tramite la quale viene poi basata la classificazione dei pazienti nelle classi “Worsening” o “Non Worsening”. Il primo obiettivo di questa tesi è stato quindi la creazione di un data frame contenente unicamente i dati alla baseline di ciascun soggetto. Con il termine “baseline” si intende qui il momento di inizio dello studio per il singolo paziente. All’atto pratico, l’intento è quindi la realizzazione di un data frame in cui ciascuna riga corrisponde a un paziente e ogni colonna rappresenta una diversa variabile associata all’individuo e ritenuta potenzialmente di interesse per la predizione dell’outcome.

Tutte le operazioni di pre-processing sono state effettuate tramite il linguaggio di programmazione R.

Dopo un’indagine approfondita dei domini appartenenti al database MSOAC Placebo, si è giunti alla decisione di escludere 5 dei 12 domini funzionali. Questi 5 domini sono:

- Dominio “CE”: il dominio “Clinical Events” raggruppa dati su relapses dei pazienti confermate (es. “confermata da MRI”) o non (es. “possibile relapse”). Tale dominio è stato escluso dall’analisi poiché tutti i dati sono riferiti a tempi successivi all’inizio dello studio (dal giorno 1 in poi);
- Dominio “RP”: il dominio “Reproductive System Findings” è un data frame contenente risultati (positivi o negativi) di test di gravidanza. Non è stato incluso nell’analisi poiché coinvolge unicamente 142 pazienti per le quali il risultato del test è sempre negativo, a meno di 3 risultati “missing”;
- Dominio “SM”: il dominio “Subject Disease Milestones” raggruppa informazioni sulle relapses confermate per 854 pazienti. Oltre a riguardare meno della metà dei pazienti, questi dati sono relativi ad attacchi che i soggetti hanno avuto nel corso del trial clinico, dal giorno 1 in poi. Quindi, come per il dominio “CE”, il dominio “SM” non include dati utili alla definizione di dati relativi al giorno 0, ovvero alla baseline degli studi;
- Dominio “DS”: il dominio “Disposition” unisce dati relativi al consenso informato dei pazienti (in tal proposito, dall’analisi di questi dati si conferma che questo è stato dato da ciascuno degli individui) a informazioni sul completamento o meno dello studio da parte dei soggetti. Dato che il presente lavoro

non si pone come obiettivo quello di una analisi di sopravvivenza o un altro tipo di studio che implichi la necessità di conoscere quest'ultima nozione, il dominio "DS" è stato scartato nella fase di pre-processing;

- Dominio "CM": il dominio "Concomitant Medications" contiene informazioni su terapie e farmaci assunte dai pazienti prima dell'ingresso nel trial clinico. 102 pazienti seguivano prima dello studio delle terapie relative alla cura della Sclerosi Multipla. Questo dominio, non è stato incluso nel lavoro della presente tesi poiché il numero di terapie e farmaci assunti dai pazienti risulta troppo elevato e senza una adeguata categorizzazione.

Il numero di domini utilizzati per la realizzazione degli obiettivi di questa tesi si riduce quindi a 7: "DM", "FA", "FT", "QS", "SC", "OE" e "MH". Per ciascuno di essi, è riportato nelle sezioni successive il lavoro di analisi effettuato. In generale, la strategia adottata allo scopo di creare un dataframe baseline può essere riassunta nei seguenti passi:

1. Studio e ispezione di ciascun dominio
2. Selezione delle variabili di interesse in ogni dominio
3. Scelta dei record alla baseline
4. Decisioni per pazienti specifici (es. nei casi in cui un paziente abbia due esiti diversi di un test alla baseline)
5. Left Join Merge sequenziale dei domini (analizzato il primo dominio, "DM", ad esso sono stati aggiunti man mano tutti gli altri domini)
6. Verifica del processo effettuato

La variabile utilizzata per il "Left Join Merge", ovvero per la sequenziale fusione dei domini analizzati, è "USUBJID", ovvero "Unique Subject Identifier", il codice univoco di identificazione di ciascun paziente, presente chiaramente per ogni record in tutto il database MSOAC Placebo. Le feature (con statistiche annesse) ricavate per il data frame di informazioni alla baseline sono riportate in tabella 2.2.

2.2.1 Dominio DM

Il dominio "DM" ("Demographics") include i dati demografici dei pazienti. All'interno di questa parte di database sono state selezionate 5 variabili di interesse:

- AGE: variabile che rappresenta l'età di ciascun paziente, espressa in anni. L'età varia tra un minimo di 18 anni e un massimo di 72. Il numero totale di NAs (Not Available data) è pari al 3.3% del totale dei pazienti. Come si denota dalla figura 2.1, la distribuzione dell'età dei soggetti è tipica di un campione di pazienti affetti da Sclerosi Multipla, con la maggior parte degli individui di età compresa tra 30 e 55 anni.

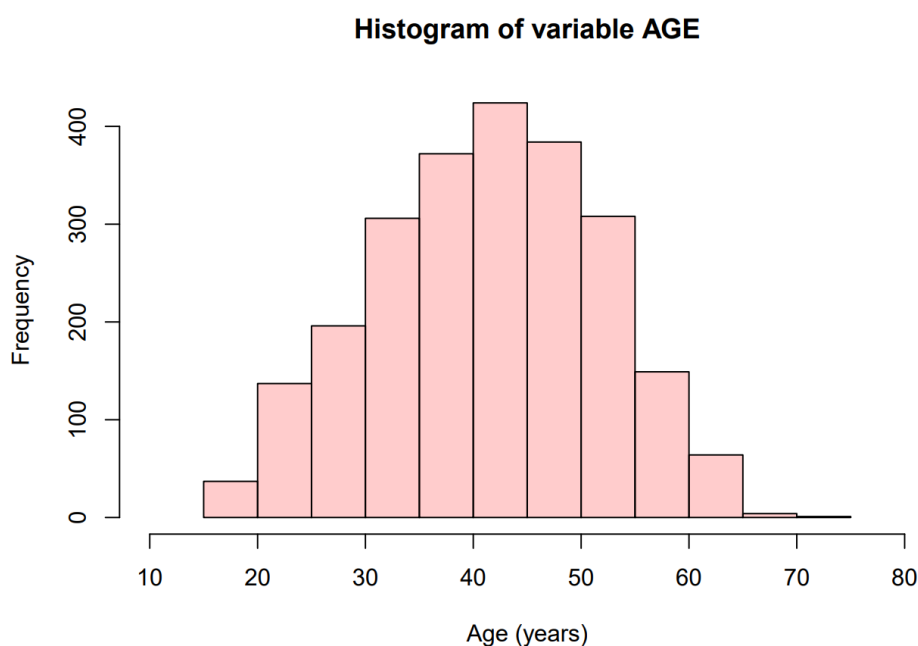


Figura 2.1: Istogramma della variabile Età

- SEX: variabile binaria che indica il sesso del paziente: “M” (Maschio) o “F” (Femmina). Le femmine sono presenti in numero maggiore (1658) rispetto ai maschi (807). Sono rispettati i rapporti tra sessi espressi in letteratura, in cui viene descritta un’incidenza della malattia superiore per il sesso femminile (rapporto in media di 2:1) [2,5]. Questa feature è presente per tutti i pazienti del database MSOAC Placebo.
- RACE: feature che riporta l’etnia di ciascun paziente. I valori, in stringhe, sono: “American Indian or Alaska native” (freq. 2/2465), “Asian” (freq. 64/2465), “Black or African American” (freq. 39/2465), “Hispanic” (freq. 10/2465), “Hispanic or Latino” (freq. 1/2465), “White” (freq. 1534/2465) e “Other” (freq. 41/2465). Il numero di dati non disponibili per la variabile

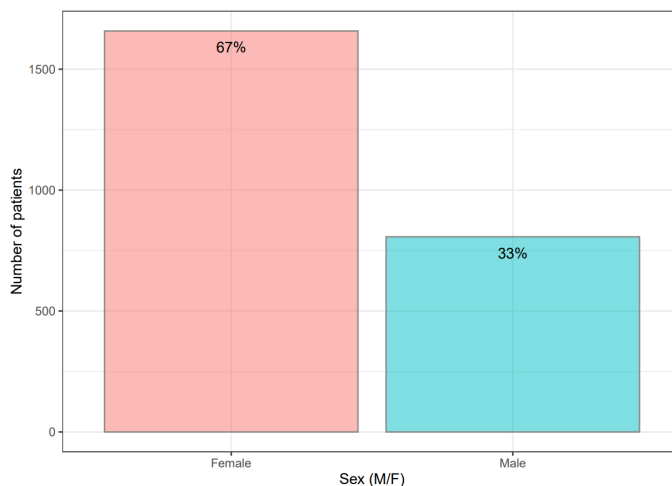


Figura 2.2: Distribuzione della variabile Sesso

“RACE” equivale al 31.5% del totale dei pazienti. Si nota che la maggior parte dei soggetti è di etnia bianca, in linea con la letteratura che riporta un’incidenza della Sclerosi Multipla superiore per questa etnia [3].

- COUNTRY: Paese di provenienza per ciascun paziente. Questa variabile è riportata in stringhe che abbreviano lo stato di nascita dei soggetti: AUS (freq. 7/2465), BEL (freq. 21/2465), BGR (freq. 21/2465), CAN (freq. 73/2464), CHE (freq. 6/2465), CHL (freq. 1/2465), COL (freq. 5/2465), CZE (freq. 63/2465), DEU (freq. 44/2465), DNK (freq. 2/2465), ESP (freq. 7/2465), EST (freq. 8/2465), FIN (freq. 3/2465), FRA (freq. 13/2465), GBR (freq. 37/2465), GEO (freq. 7/2465), GRC (freq. 14/2465), HRV (freq. 5/2465), HUN (freq. 19/2465), IND (freq. 56/2465), IRL (freq. 3/2465), ISR (freq. 6/2465), LVA (freq. 3/2465), MEX (freq. 9/2465), NLD (freq. 26/2465), NZL (freq. 10/2465), PER (freq. 7/2465), POL (freq. 177/2465), ROU (freq. 16/2465), RUS (freq. 48/2465), SRB (freq. 46/2465), SWE (freq. 9/2465), TUR (freq. 5/2465), UKR (freq. 63/2465) e USA (freq. 249/2465). Ispezionando quali paesi hanno frequenza maggiore, anche qui, si denota che il database rispecchia bene la statistica in letteratura per la distribuzione geografica della malattia. Difatti la maggioranza dei pazienti proviene da paesi distanti dall’equatore, dove sembra esserci una incidenza più rilevante. Nel caso del database MSOAC Placebo, molti soggetti provengono da paesi dell’emisfero nord, come Polonia, Stati Uniti d’America e Canada.
- ETHNIC: Etnia dei soggetti riportata come “Hispanic or Latino” o “Not Hispanic or Latino”. La prima è presente nel database per 11 pazienti, la seconda

per 217 (91% di NAs per questa variabile nel database MSOAC Placebo).

2.2.2 Dominio FA

Il dominio “FA” contiene dati riguardanti il numero di ricadute (relapses) confermate per ciascun paziente nei tre anni precedenti all’inizio dello studio o dalla diagnosi della malattia (eseguita tramite il Criterio di Mc Donald [17]). Da questo dominio sono state estratte quattro variabili:

- P1Y: Numero di relapses che il soggetto ha subito nell’anno precedente all’inizio dello studio;
- P2Y : Numero di relapses che il paziente ha avuto due anni prima dell’inizio dello studio;
- P3Y: Numero di relapses che il soggetto ha sperimentato tre anni prima dell’ingresso nello studio;
- S_MS_D: Numero di relapses del paziente tra la diagnosi della malattia e l’inizio dello studio.

2.2.3 Dominio FT

Il dominio “FT” è risultato tra i più interessanti per questo lavoro di tesi. In questa parte del database si trovano dati raccolti su test funzionali svolti dai pazienti. I test verificano sia le capacità fisiche degli individui (es. mobilità degli arti superiori o inferiori) sia le prestazioni cognitive. Le feature ricavate dal dominio “FT” riguardano i seguenti test:

- Test T25FW: il “Time 25 Foot Walk Test” è definito come “la miglior misura oggettiva per descrivere la disabilità di camminata per pazienti affetti da Sclerosi Multipla che può essere utilizzata in un ampio range di disabilità degli individui in esame” [44]. Il T25FW è un test di camminata per il quale viene richiesto al paziente di camminare il più velocemente possibile ma in sicurezza un percorso rettilineo di lunghezza pari a 25 piedi (7.62 metri). Il risultato è espresso in “secondi impiegati per concludere il tragitto”; è un test pratico che richiede un minimo spazio ed equipaggiamento per la sua realizzazione e di conseguenza ha un basso costo di attuazione. Inoltre, è stata dimostrata la sua affidabilità nel breve e lungo termine per valutare pazienti con EDSS compreso tra 0 e 6.5, il che lo rende una misura significativa dal punto di vista

clinico [45]. Oltre a ciò, essendo un test di “camminata breve”, è rilevante per la vita quotidiana dei pazienti, per esempio nei casi di attraversamento pedonale, rispondere al telefono o andare al bagno [45]. Attualmente, una variazione del 20% nel tempo del risultato del test viene considerata un cambiamento rilevante nelle performance di camminata di un paziente affetto da MS [45]. Alla baseline di un trial clinico vengono effettuati solitamente due test T25FW consecutivi. Vengono poi mediati i due risultati ottenendo il valore finale per ciascun paziente. In questo lavoro di tesi sono stati salvati dal dominio “FT” i due risultati del test T25FW ad inizio studio per ottenere le variabili “T25FW_1” e “T25FW_2”. E’ stata inoltre presa in considerazione un’ulteriore variabile relativa a questo test, presente nel database: T25FW_MT2A, dove “MT2A” è l’acronimo di “More Than 2 Attempts”. Questa variabile, espressa in “Y” (Yes) o “N” (No), indica se sono stati necessari più di due tentativi per concludere i test consecutivi T25FW alla baseline.

- Test SDMT: il “Symbol Digit Modality Test” è un test cognitivo che individua il rallentamento della velocità di elaborazione delle informazioni in pazienti malati di Sclerosi Multipla, uno tra i sintomi più comuni della patologia. Per lo svolgimento del SDMT viene presentato al paziente un foglio in cui sono presenti 9 numeri ai quali sono associati 9 simboli. Successivamente, al soggetto viene richiesto di associare a ciascun simbolo presentato il numero relativo ad esso; il risultato del test è riportato come “numero di associazioni corrette che il paziente svolge in 90 secondi”. Ad oggi, il SDMT viene eseguito principalmente nella sua versione orale, per non avere influenze dalla capacità motoria dell’individuo [46]. Il test ha presentato inoltre una forte correlazione con varie metriche di “processing speed” ottenute tramite MRI. All’interno del dominio “FT” è stata ricavata la feature “SDMT”, la quale riporta il risultato dell’omonimo test eseguito dai pazienti.
- Test PASAT: il “Paced Auditory Serial Additional Task” è un test cognitivo utilizzato per indagare la velocità e la flessibilità di elaborazione delle informazioni uditive, oltre all’abilità di calcolo [47]. Al paziente vengono presentati oralmente numeri in sequenza; l’obiettivo è quello di sommare ciascun numero con i precedenti. I numeri possono essere presentati ogni 3 secondi (PASAT-3s), secondo la versione classica, oppure a intervalli di 2 secondi (PASAT-2s), per cui viene aumentata notevolmente la difficoltà del test. Il risultato del test viene riportato come “numero di risposte corrette eseguite in 60 possibili totali”. Questo test si è dimostrato affidabile, ad alta sensibilità e con buo-

ne proprietà psicometriche [48]. Nel database MSOAC Placebo, alla baseline, sono presenti entrambe le tipologie del test: PASAT-3s e PASAT-2s, per le quali sono state create due variabili. Inoltre, è stata estratta anche la feature PASAT_MT1A: questa variabile indica tramite le stringhe “Y” (Yes) o “N” (No) se è stato necessario più di un tentativo affinché il paziente sia riuscito a svolgere il test nella versione tradizionale (3 secondi).

- Test NHPT: il “Nine Hole Peg Test” è un test rivolto all’analisi della capacità motoria degli arti superiori, in particolare è utile nel valutare la destrezza manuale dei pazienti affetti da Sclerosi Multipla; ad oggi è considerato il “gold standard” per questo scopo. Il NHPT richiede al soggetto in esame di spostare e poi rimuovere il più velocemente possibile 9 pioli in 9 buchi corrispondenti (uno alla volta). Il risultato è espresso in secondi necessari per concludere il test. Circa il 53% della varianza del risultato è spiegata dalla forza muscolare, sensibilità tattile del pollice e presenza di tremore [49]. Un cambiamento del 20% nel tempo del risultato del test suggerisce un peggioramento del paziente affetto dalla malattia [50]. Nel database MSOAC Placebo, alla baseline, sono presenti 4 NHPT: due eseguiti consecutivamente con la mano dominante e due, anch’essi eseguiti consecutivamente, con la mano non dominante. Per ciascuno di essi si è ricavata una feature, per un totale di 4 feature relative al test NHPT. A partire dall’informazione a disposizione, si sono inoltre codificate altre due feature del tipo More Than 2 Attempts (MT2A, analoghe a quelle presenti per il Test T25FW), che riportano se sono stati necessari più di due tentativi per concludere i test con mano dominante e non.

2.2.4 Dominio QS

Il dominio “QS” (Questionnaires) include i risultati di questionari proposti ai pazienti o eseguiti dal personale clinico. Per ciascuno sono state considerate, come in ogni dominio, unicamente le risposte salvate alla baseline. In generale, i questionari selezionati si possono suddividere in 5 categorie, dalle quali sono state poi estratte le variabili di interesse:

- Questionario BDI: il “Beck Depression Inventory” è un questionario proposto da Aaron T. Beck per valutare la severità della depressione per pazienti malati di Sclerosi Multipla [51]. Il BDI è composto da 21 quesiti che quantificano problemi cognitivi e fisici relazionati alla depressione che l’individuo ha sperimentato nella settimana precedente allo svolgimento del questionario. Ogni

domanda ha 4 possibili risposte, alle quali è associato un risultato numerico nel range 0-3. Maggiore è il numero corrispondente alla risposta, maggiore è la negatività della risposta; per esempio, nel primo quesito, rivolto al grado di tristezza per l'individuo, le risposte spaziano da “non mi sento triste”, score di 0, a “sono così triste o infelice che non riesco più a sopportarlo” corrispondente allo score 3. Nel database MSOAC Placebo sono riportati i risultati per ciascuna delle 21 domande, oltre ai quali è presente il risultato complessivo dato dalla somma delle 21 risposte. Per ciascuno dei quesiti (incluso lo score totale) sono state create delle feature per il data frame baseline.

- Score EDSS: come introdotto nella sezione 1.8, la scala EDSS riporta un valore compreso tra 0 e 10 descrittivo della disabilità del paziente. Maggiori sono le disabilità accumulate dal paziente, maggiore è il valore di EDSS ad egli associato. In figura 2.3 viene riportata la distribuzione del valore EDSS alla baseline nei dati, dalla quale si nota una buona eterogeneità dei questo score nei pazienti all'inizio dello studio. Il range di valori va da 0 ad un massimo di 6.5. La moda si stabilisce sul valore 2; in media, invece, i pazienti hanno un valore EDSS alla baseline pari a 3.4.

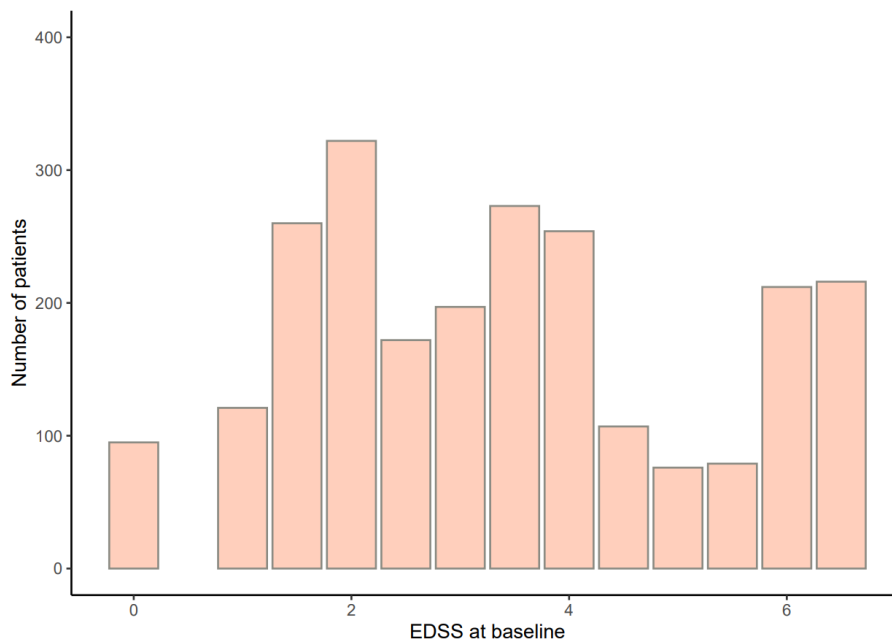


Figura 2.3: Distribuzione del valore EDSS alla baseline

- Questionario KFSS: il “Kurtzke Functional Systems Scores” (KFSS) è un questionario compilato dal personale clinico che riporta in un range variabile tra 0

e 9 il risultato di disabilità dei pazienti su sette sistemi funzionali: funzioni piramidali, funzioni cerebellari, funzioni del tronco cerebrale, funzioni sensoriali, funzione intestinale e vescicale, funzioni visive e funzioni cerebrali (o mentali) [52]. Per ciascun sistema funzionale, più importante è la disabilità del paziente e maggiore è il risultato dello score KFSS corrispondente. Dal dominio “QS”, è stata ricavata una feature per ciascun quesito KFSS.

- **Questionario RAND-36:** il RAND-36 è un questionario creato dall’organizzazione di ricerca RAND Corporation. Questo strumento è il più utilizzato al mondo per la valutazione della Health Related Quality of Life (HRQoF) dei pazienti, ovvero la qualità di vita relazionata alla salute [53]. Il RAND-36 si compone di 36 quesiti che rilevano il funzionamento fisico, il dolore fisico, limitazioni dovute alla salute fisica, limitazioni dovute a problemi personali o emotivi, salute mentale generale, funzionamento sociale, energia/stanchezza e percezioni generali di salute [54]. In questo lavoro di tesi sono state ricavate 36 feature alla baseline, ciascuna relativa al risultato dei 36 quesiti del questionario RAND-36.
- **Questionario SF12:** il “Short Form 12” è la versione ridotta (come numero di quesiti ma mantenendo tutti i domini di indagine) del questionario Short Form Health Survey 36 (SF-36). Il SF12 è incentrato sulla salute del soggetto che lo compila, e studia 8 differenti dimensioni: percezione generale della salute, salute fisica, funzione di ruolo fisico limitata, dolore fisico, vitalità, salute mentale, funzione di ruolo emotivo limitata e funzionamento sociale [55]. Il punteggio di ciascuna domanda del questionario varia nell’intervallo 1-5. Maggiore è la negatività della risposta maggiore è il punteggio nel quesito associato. In totale, 12 feature sono ricavate dalle risposte al questionario SF12, ciascuna relativa al punteggio stabilito dall’individuo in ogni quesito.

2.2.5 Dominio SC

Il dominio “SC” (Subject Characteristics) include come unico dato utile la mano dominante di ciascun individuo del database MSOAC Placebo. Essendo il peggioramento della destrezza manuale uno dei principali sintomi della Sclerosi Multipla, la mano dominante dei soggetti affetti dalla malattia può cambiare nel tempo. Per questo dominio è stata estratta per cui la variabile “DH”, ovvero Dominant Hand, mano dominante del paziente all’inizio dello studio, riportata come “L” (Left) se il soggetto è mancino alla baseline oppure “R” (Right) se risulta essere destro.

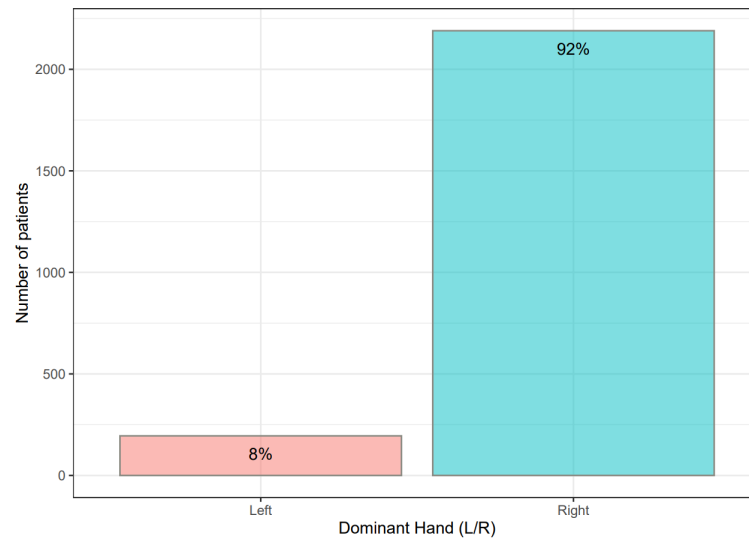


Figura 2.4: Distribuzione mano dominante (DH) alla baseline

2.2.6 Dominio OE

Il dominio “OE” (Ophtalmic Examinations) include informazioni riguardanti test visivi svolti dai pazienti. La disfunzione del sistema visivo è uno tra i sintomi più frequenti della Sclerosi Multipla. Ad oggi, esistono svariate misure della disabilità visiva; nel database MSOAC Placebo, dal dominio “OE”, sono state ricavate 4 feature, tutte basate su test di acuità visiva:

- Snelleq: lo “Snellen Equivalent Score” è una misura di acuità visiva ottenuta dividendo: la distanza tra soggetto in esame e la “Snellen Chart”, ovvero la tavola visiva contenente lettere di diverse dimensioni, e la distanza alla quale il più piccolo ottotipo identificato sottende un angolo di cinque minuti d’arco. Il risultato è per cui espresso in metri (m).
- SLEC 100%: lo “Sloan Letter Eye Chart 100%” è un test di acuità visiva ad alto contrasto, per il quale le lettere da visualizzare per il paziente sono nere su uno sfondo bianco. Questo test, nonostante la sua importanza in numerosi ambiti, è talvolta criticato perché ritenuto essere uno strumento non sufficientemente preciso nel distinguere pazienti sani e pazienti affetti da Sclerosi Multipla [56].
- SLEC 1.25% e SLEC 2.5%: gli “Sloan Letter Eye Chart” a 1.25% e 2.5% sono test visivi che indagano l’acuità visiva a basso contrasto. Le percentuali 1.25% e 2.5% si riferiscono alla percentuale di contrasto delle lettere rispetto al

100%. Più bassa è la percentuale, più le lettere assumono un grigio maggiormente chiaro su sfondo bianco. Questi due SLEC test risultano fondamentali, poiché è stato dimostrato che pazienti affetti da Sclerosi Multipla si distinguono notevolmente da soggetti sani per quanto riguarda il risultato di test di acuità visiva a basso contrasto (risultati molto inferiori per individui MS rispetto a soggetti di controllo) [56]. Inoltre, questi deficit sono rilevati anche anni dopo un episodio di neurite ottica acuta, quando invece l'acuità ad alto contrasto è completamente recuperata da parte del paziente [56].

2.2.7 Dominio MH

Il dominio “MH” (Medical History) riporta dati riguardanti la diagnosi dei pazienti e delle preesistenti condizioni mediche. Da questo dominio è stata estratta l'indispensabile nozione del fenotipo di ciascun soggetto al momento dell'inizio dello studio. I pazienti del database MSOAC Placebo hanno alla baseline una tra le forme RRMS, SPMS, PPMS e PRMS. Si segnala che l'ultima forma, PRMS (Progressive Relapsing Multiple Sclerosis), ad oggi non viene più considerata un vero e proprio fenotipo. Questa forma è segnata nel database per soli due pazienti, che sono pertanto stati esclusi dall'analisi di questo lavoro, riducendo così la cardinalità del dataset da 2465 a 2463 soggetti. Nel database MSOAC Placebo il dato relativo al fenotipo di MS per ciascun soggetto è presente solo all'inizio dello studio. Di conseguenza, non è possibile studiare la conversione da fenotipo RRMS a SPMS. Ciò ha portato all'esclusione di quest'ultima come outcome dei modelli successivamente creati.

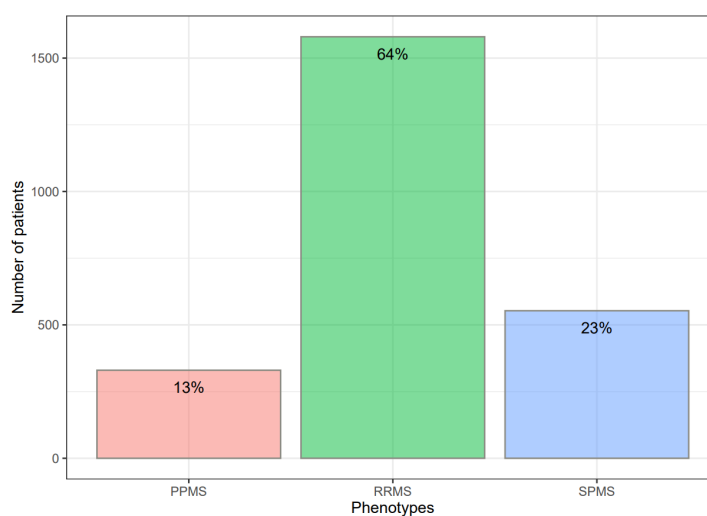


Figura 2.5: Distribuzione dei fenotipi alla baseline

2.3 Feature Selection by NAs e andamento temporale del parametro EDSS

Il dataframe di dati alla baseline dei pazienti risulta quindi una matrice con 2463 righe (pari al numero totale di pazienti del database MSOAC Placebo escludendo i due pazienti con fenotipo PRMS) e 104 colonne, corrispondenti alle feature ricavate secondo quanto illustrato nella sezione 2.2. Per ciascuna variabile (o covariata) è stata quindi effettuata una scelta di mantenimento o eliminazione basata sulla percentuale di NAs (Not Available Data). La soglia per la selezione è stata posta al 30%: le variabili con più del 30% di valori mancanti sono state rimosse dal data frame alla baseline. A seguito di questo filtraggio, il numero di feature considerato si è ridotto a un totale di 23. Dopo questa prima scrematura, sono state scartate anche le feature considerate poco utili ai fini dei modelli predittivi, poiché portatrici di poche informazioni. Queste feature sono: T25FW_MT2A, NHPT_DH_MT2A e NHPT_NDH_MT2A. Esse difatti vengono espresse in “Y” (Yes) o “N” (No), ma, per ciascuna, la sproporzione tra i risultati è troppo elevata (risposte Y in numero troppo basso per poter essere rilevanti). Il numero di feature totale giunge infine a 20, elencate qui di seguito assieme al dominio di appartenenza (riportato per ciascuna tra parentesi):

- AGE (DM)
- SEX (DM)
- P1Y (FA)
- T25FW_1 (FT)
- T25FW_2 (FT)
- PASAT (FT)
- NHPT_DH_1 (FT)
- NHPT_DH_2 (FT)
- NHPT_NDH_1 (FT)
- NHPT_NDH_2 (FT)
- EDSS (QS)
- KFSS1 (QS)

- KFSS2 (QS)
- KFSS3 (QS)
- KFSS4 (QS)
- KFSS5 (QS)
- KFSS6 (QS)
- KFSS7 (QS)
- DH (SC)
- FENOTIPO (MH)

La fase successiva alla creazione del dataframe baseline è stata la scelta dell'outcome da predire tramite i modelli computazionali successivamente sviluppati. Come già accennato nella sezione 2.2.7, nel database MSOAC Placebo non sono presenti dati riguardanti l'evoluzione (o il tempo per impiegato per essa) dal fenotipo RRMS a quello SPMS, il che ha portato all'esclusione di questo outcome frequentemente impiegato in letteratura. Tra le ipotesi vagliate, si è considerata la possibilità di selezionare come outcome il peggioramento della malattia in base al test T25FW o al test NHPT, definito, in entrambi i casi, come una variazione di almeno il 20%. Queste opzioni sono tuttavia state escluse perché raramente trovate nello stato dell'arte attuale. Se ne propone comunque un possibile impiego in studi successivi a questo lavoro di tesi.

A seguito di queste considerazioni ed analisi, l'outcome scelto per questo lavoro di tesi corrisponde a quello maggiormente utilizzato negli studi di letteratura, ovvero il Δ EDSS che, come introdotto nella sezione 1.8, corrisponde alla differenza tra il valore EDSS in un certo momento futuro ed il valore dello stesso alla baseline. Si è quindi proceduto a ispezionare l'andamento della variabile EDSS nel tempo, per capire per quanti soggetti è presente il dato in diversi istanti temporali, al fine di determinare la soglia a cui calcolare il Δ EDSS. In figura 2.6 è riportata graficamente la numerosità dei soggetti che presentano una rilevazione EDSS a diversi istanti temporali (in un intervallo compreso tra la baseline e 36 mesi dall'inizio dello studio, a passo di 3 mesi).

Da questa analisi si può osservare come il numero di pazienti che dispongono del dato EDSS diminuisca nel tempo. Come emerge dal grafico, più di metà dei pazienti (totale di 2463) hanno almeno un rilevazione del valore EDSS fino a 18-24 mesi. Oltre i 24 mesi, il numero di individui che dispongono del valore EDSS diminuisce

drasticamente, fino ad arrivare a 3 anni, per cui si ha a disposizione il valore del parametro per all'incirca 600 soggetti. Dato che uno dei punti di forza dello studio oggetto di questa tesi consiste nella possibilità di avere un numero di soggetti superiore alla media riscontrata in letteratura (vedi sezione 1.8), si è deciso di fissare una soglia temporale pari a 18 mesi per il calcolo del Δ EDSS, in modo da mantenere una cardinalità dei soggetti consistente.

Selezionati l'outcome e l'orizzonte temporale di previsione, questo lavoro è incentrato nello sviluppo di modelli predittivi di classificazione. Nel definire le classi, si è utilizzata la label "Worsening" per contrassegnare i pazienti con Δ EDSS maggiore di una determinata soglia, la label "Non Worsening" viene applicata ad un soggetto per cui invece il Δ EDSS è inferiore alla soglia. Le soglie utilizzate per la classificazione dei pazienti verranno sintetizzate nel successivo Capitolo 3, nella sezione 3.3. L'algoritmo scelto per allenare i diversi modelli è stato la Regressione Logistica. Nel prossimo capitolo, si descrive il processo di implementazione dei modelli, svolto nel linguaggio di programmazione R, così come tutta la fase di pre-processing.

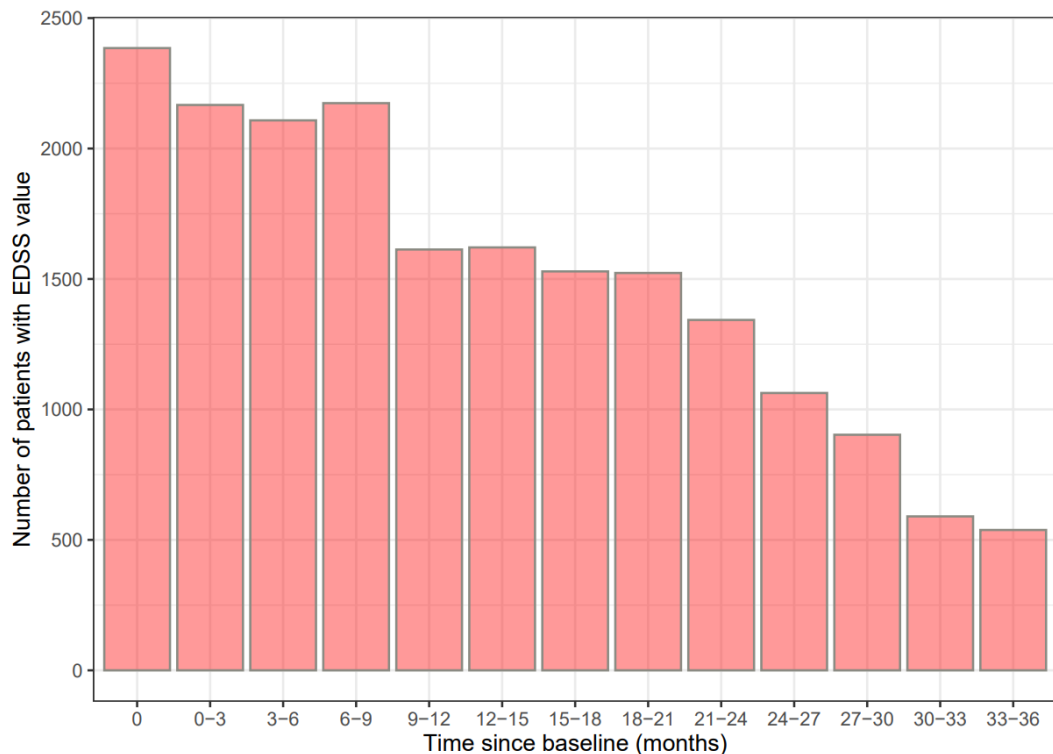


Figura 2.6: Numero di pazienti che hanno il dato EDSS da baseline a 36 mesi, ad intervalli di 3 mesi in 3 mesi

Tabella 2.2: Statistiche e variabili alla baseline

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
USUBJID	UNIQUE SUBJECT IDENTIFIER	CHARACTER	-	0%	-	-
Dominio DM						
AGE	ETA'	CHARACTER	YEARS	3.37%	[18,72]	41.77
SEX	SESSO	CHARACTER	M/F	0%	-	-
RACE	ETNIA	CHARACTER	-	31.4%	-	-
COUNTRY	PAESE	CHARACTER	-	56,67%	-	-
ETHNIC	LATINO/ISPANICO o NO	CHARACTER	-	90.75%	-	-
Dominio FA						
P1Y-RELAPSES	NUMERO DI RICADUTE DEL PAZIENTE NELL'ANNO -1	NUMERIC	-	27.75%	[0, 7]	1.376
P2Y-RELAPSES	NUMERO DI RICADUTE DEL PAZIENTE NELL'ANNO -2	NUMERIC	-	68.68%	[1, 14]	2.18
P3Y-RELAPSES	NUMERO DI RICADUTE DEL PAZIENTE NELL'ANNO -3	NUMERIC	-	59.11%	[0, 18]	2.336
SD_RELAPSES	Numero di ricadute del paziente dalla Diagnosi	NUMERIC	-	69.05%	[0, 50]	4.569
Dominio FT						
T25.FW.first	Primo test T25FW effettuato: tempo impiegato a percorrere un tratto rettilineo di 25 piedi.	NUMERIC	SEC	0.81%	[0.7, 218.2]	9.217
T25FW.second	Secondo test T25FW effettuato alla baseline.	NUMERIC	SEC	0.85%	[0.7, 232.1]	9.095
T25FW_MT2A	Indice che indica se sono stati necessari più di due tentativi per completare due trial di test T25FW	CHARACTER	Y/N	13.39%	-	-
SDMT01	Symbol Digit Modalities Test, per performance cognitive (il risultato è il numero di associazioni corrette in 90 sec)	NUMERIC	-	79.76%	[0, 110]	48.68
PASAT_tc.3s	PASAT-total correct 3 secondi: Paced Auditory Serial Additional Task, test sull'attenzione utitiva/verbale. I numeri presentati oralmente vengono dati ogni 3 sec. Risultato in "n° risposte corrette su 60 totali"	NUMERIC	-	3.41%	[0, 60]	47.24
PASAT_tc.2s	PASAT-total correct 2 secondi: medesimo al test precedente ma a ritmo di 2 secondi, molta più difficoltà	NUMERIC	-	69.57%	[3, 60]	37.01

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio FT						
PASAT_MT1A	PASAT more than one attempt: se necessario più di un tentativo per concludere il test 3s	NUMERIC	Y/N	46.21%	-	-
NHPT01_DH1	Primo Nine Hole Peg Test con mano dominante. Test sulla mobilità della mano (estensione arto scelto). Spostare pioli da un contenitore, risultato in tempo impiegato per completare il test	NUMERIC	SEC	3.41%	[1, 300]	25.1
NHPT01_DH2	Secondo NHPT <i>Defettuato alla baseline</i>	NUMERIC	SEC	3.45%	[8.5, 300]	23.88
NHPT01_NDH1	Pimo NHPT effettuato con la mano non dominante	NUMERIC	SEC	3.77%	[1.4, 312.7]	26.87
NHPT01_NDH2	Secondo NHPT con mano non dominante	NUMERIC	SEC	3.85%	[1.2, 356.7]	26.16
NHPT2A_DH	Se sono serviti più di due tentativi per completare due successful trials con mano dominante per il test NHPT01	CHARACTER	Y/N	27.71%	-	-
NHPT2A_NDH	Se sono serviti più di due tentativi per il test NHPT01 con mano non dominante	CHARACTER	Y/N	27.91%	-	-
Dominio QS						
BDI0101	SADNESS: 0 = I do not feel sad; 1 = I feel sad much of the time; 2 = I am sad all the time; 3 = I am so sad or unhappy that I can't stand it	NUMERIC	-	72.29%	[0, 3]	0.2972
BDI0102	PESSIMISM: 0 = I am not discouraged about my future; 1 = I feel more discouraged about my future than I used to; 2 = I do not expect things to work out for me; 3 = I feel my future is hopeless and will only get worse	NUMERIC	-	72.29%	[0, 3]	0.552
BDI0103	PAST FAILURE 0 = I do not feel like a failure; 1 = I have failed more than I should have; 2 = As I look back, I see a lot of failures; 3 = I feel I am a total failure as a person	NUMERIC	-	72.29%	[0, 3]	0.3631

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio QS						
BDI0104	LOSS OF PLEASURE 0 = I get as much pleasure as I ever did from the things I enjoy; 1 = I don't enjoy things as much as I used to; 2 = I get very little pleasure from the things I used to enjoy; 3 = I can't get any pleasure from the things I used to enjoy.	NUMERIC	-	72.29%	[0, 3]	0.5681
BDI0105	GUILTY FEELINGS 0 = I don't feel particularly guilty; 1 = I feel guilty over many things I have done or should have done; 2 = I feel quite guilty most of the time; 3 = I feel guilty all of the time	NUMERIC	-	72.33%	[0, 3]	0.3299
BDI0106	PUNISHMENT FEELINGS 0 = I don't feel I am being punished; 1 = I feel I may be punished; 2 = I expect to be punished; 3 = I feel I am being punished	NUMERIC	-	72.33%	[0, 3]	0.4076
BDI0107	SELF DISLIKE 0 = I feel the same about myself as ever; 1 = I have lost confidence in myself; 2 = I am disappointed in myself; 3 = I dislike myself	NUMERIC	-	72.29%	[0, 3]	0.4114
BDI0108	SELF-CRITICALNESS 0 = I don't criticize or blame myself more than usual; 1 = I am more critical of myself than I used to be; 2 = I criticize myself for all of my faults; 3 = I blame myself for everything bad that happens	NUMERIC	-	72.33%	[0, 3]	0.4707
BDI0109	SUICIDIAL THOUGHTS OR WISHES 0 = I don't have any thoughts of killing myself; 1 = I have thoughts of killing myself, but I would not carry them out; 2 = I would like to kill myself; 3 = I would kill myself if I had the chance	NUMERIC	-	72.33%	[0, 3]	0.1276
BDI0110	CRYING 0 = I don't cry anymore than I used to; 1 = I cry more than I used to; 2 = I cry over every little thing; 3 = I feel like crying, but I can't	NUMERIC	-	72.29%	[0, 3]	0.4817

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio QS						
BDI0111	AGITATION 0 = I am no more restless or wound up than usual; 1 = I feel more restless or wound up than usual; 2 = I am so restless or agitated that it's hard to stay still; 3 = I am so restless or agitated that I have to keep moving or doing	NUMERIC	-	72.45%	[0, 3]	0.514
BDI0112	LOSS OF INTEREST 0 = I have not lost interest in other people or activities; 1 = I am less interested in other people or things than before; 2 = I have lost most of my interest in other people or things; 3 = It's hard to get interested in anything	NUMERIC	-	72.49%	[0, 3]	0.4027
BDI0113	INDECIVINESS 0 = I make decisions about as well as ever; 1 = I find it more difficult to make decisions than usual; 2 = I have much greater difficulty in making decisions than I used to; 3 = I have trouble making any decisions	NUMERIC	-	72.45%	[0, 3]	0.4904
BDI0114	WORTHLESSNESS 0 = I do not feel I am worthless; 1 = I don't consider myself as worthwhile and useful as I used to; 2 = I feel more worthless as compared to other people; 3 = I feel utterly worthless	NUMERIC	-	72.45%	[0, 3]	0.3859
BDI0115	LOSS OF ENERGY 0 = I have as much energy as ever; 1 = I have less energy than I used to have; 2 = I don't have enough energy to do very much; 3 = I don't have enough energy to do anything	NUMERIC	-	72.45%	[0, 3]	0.9514
BDI0116	CHANGING IN SLEEPING PATTERN 0 = I have not experienced any change in my sleeping pattern; 1 = I sleep somewhat less than usual; 2 = I sleep a lot less than usual; 3 = I wake up 1-2 hours early and can't get back to sleep	NUMERIC	-	72.54%	[0, 3]	0.808

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio QS						
BDI0117	IRRITABILITY 0 = I am no more irritable than usual; 1 = I am more irritable than usual; 2 = I am much more irritable than usual; 3 = I am irritable all the time	NUMERIC	-	72.45%	[0, 3]	0.5405
BDI0118	CHANGES IN APPE- TITE 0 = I have not experienced any change in my appetite; 1 = My appetite is somewhat greater than usual/My appetite is somewhat less than usual; 2 = My appetite is much less than before./My appetite is much greater than usual; 3 = I crave food all the time/I have no appetite at all	NUMERIC	-	72.45%	[0, 3]	0.458
BDI0119	CONCENTRATION DIFFICULTY 0 = I can concentrate as well as ever; 1 = I can't concentrate as well as usual; 2 = It's hard to keep my mind on anything for very long; 3 = I find I can't concentrate on anything	NUMERIC	-	72.49%	[0, 3]	0.6563
BDI0120	TIREDFNESS OR FA- TIGUE 0 = I am no more tired or fatigued than usual; 1 = I get more tired or fatigued more easily than usual; 2 = I am too tired or fatigued to do a lot of the things I used to do; 3 = I am too tired or fatigued to do most of the things I used to do	NUMERIC	-	72.45%	[0, 3]	0.9588
BDI0121	LOSS OF INTEREST IN SEX 0 = I have not noticed any recent change in my interest in sex; 1 = I am less interested in sex than I used to be; 2 = I am much less interested in sex now; 3 = I have lost interest in sex completely	NUMERIC	-	72.45%	[0, 3]	0.5552
BDI0122TOT	SCORE TOTALE rea- lizzato come somma dei risultati per ciascun paziente dai 21 test BDI01 precedenti	NUMERIC	-	72.37%	[0, 52]	10.75

CAPITOLO 2. IL DATABASE MSOAC PLACEBO: ANALISI E
PRE-PROCESSING

46

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio QS						
EDSS	Scala EDSS su invalidità del paziente (0-10)	NUMERIC	-	3.29%	[0, 6.5]	3.406
KFSS101	KFSS101: score su PYRAMIDAL FUNCTIONS	NUMERIC	-	15.62%	[0, 5]	1.853
KFSS102	KFSS102: score su CEREBELLAR FUNCTIONS	NUMERIC	-	16.35%	[0, 9]	1.29
KFSS103	KFSS103: score su BRAIN STEM FUNCTIONS	NUMERIC	-	15.62%	[0, 4]	0.6351
KFSS104	KFSS104: score su SENSORY FUNCTIONS	NUMERIC	-	15.62%	[0, 5]	1.178
KFSS105	KFSS105: score su BOWEL AND BLADDER FUNCTIONS	NUMERIC	-	15.62%	[0, 6]	0.8322
KFSS106	KFSS106: score su VISUAL OR OPTIC FUNCTIONS	NUMERIC	-	15.62%	[0, 6]	0.7543
KFSS107	KFSS107: score su CEREBRAL OR MENTAL FUNCTIONS	NUMERIC	-	15.62%	[0, 3]	0.5827
R360101	R360101: Would You Say Your Health Is (>=WORSE)	NUMERIC	-	57.61%	[1, 5]	3.028
R360102	R360102: Health Now Compared to 1 Year Ago	NUMERIC	-	57.24%	[1, 5]	3.458
R360103	R360103: Health Limit Vigorous Activities	NUMERIC	-	57.2%	[1, 3]	1.272
R360104	R360104: Health Limit Moderate Activities	NUMERIC	-	57.16%	[1, 3]	1.782
R360105	R360105: Health Limit Carrying Groceries	NUMERIC	-	57.48%	[1, 3]	1.946
R360106	R360106: Health Limit Climb Several Flights	NUMERIC	-	57.28%	[1, 3]	1.595
R360107	R360107: Health Limit Climb One Flight	NUMERIC	-	57.28%	[1, 3]	2.115
R360108	R360108: Health Limit Bending or Kneeling	NUMERIC	-	57.24%	[1, 3]	1.992
R360109	R360109: Health Limit Walk More Than Mile	NUMERIC	-	57.28%	[1, 3]	1.47
R360110	R360110: Health Limit Walk Several Blocks	NUMERIC	-	57.28%	[1, 3]	1.712
R360111	R360111: Health Limit Walk One Block	NUMERIC	-	57.2%	[1, 3]	2.097
R360112	R360112: Health Limit Bathing/Dressing Self	NUMERIC	-	57.16%	[1, 3]	2.516
R360113	R360113: Phys Health: Cut Work Time	NUMERIC	-	57.36%	[1, 2]	1.54
R360114	R360114: Phys Health: Accomplish Less	NUMERIC	-	57.44%	[1, 2]	1.346
R360115	R360115: Phys Health: Limit Kind of Work	NUMERIC	-	57.36%	[1, 2]	1.368
R360116	R360116: Phys Health: Difficulty Performing	NUMERIC	-	57.24%	[1, 2]	1.358
R360117	R360117: Emotional Problem: Cut Work Time	NUMERIC	-	57.28%	[1, 2]	1.724

2.3. FEATURE SELECTION BY NAS E ANDAMENTO TEMPORALE DEL PARAMETRO EDSS

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio QS						
R360117	R360117: Emotional Problem: Cut Work Time	NUMERIC	-	57.28%	[1, 2]	1.724
R360118	R360118: Emotional Problem: Accomplish Less	NUMERIC	-	57.24%	[1, 2]	1.606
R360119	R360119: Emotional Problem: Less Careful	NUMERIC	-	57.36%	[1, 2]	1.732
R360120	R360120: Extent Interfered Social Activity	NUMERIC	-	57.36%	[1, 5]	2.193
R360121	R360121: How Much Bodily Pain Have You Had	NUMERIC	-	57.24%	[1, 6]	2.62
R360122	R360122: Did Pain Interfere With Your Work	NUMERIC	-	57.48%	[1, 5]	2.004
R360123	R360123: Did You Feel Full of Pep	NUMERIC	-	57.48%	[1, 6]	3.913
R360124	R360124: Have Been a Very Nervous Person	NUMERIC	-	57.28%	[1, 6]	4.744
R360125	R360125: Have Felt So Down in the Dumps	NUMERIC	-	57.36%	[1, 6]	5.224
R360126	R360126: Have You Felt Calm and Peaceful	NUMERIC	-	57.44%	[1, 6]	3.114
R360127	R360127: Did You Have a Lot of Energy	NUMERIC	-	57.44%	[1, 6]	4.139
R360128	R360128: Have Felt Downhearted and Blue	NUMERIC	-	57.36%	[1, 6]	4.817
R360129	R360129: Did You Feel Worn Out	NUMERIC	-	57.32%	[1, 6]	3.692
R360130	R360130: Have You Been a Happy Person	NUMERIC	-	57.53%	[1, 6]	2.807
R360131	R360131: Did You Feel Tired	NUMERIC	-	57.36%	[1, 6]	3.224
R360132	R360132: Amt Time Interfere Social Activity	NUMERIC	-	57.48%	[1, 5]	3.697
R360133	R360133: Get Sick Easier Than Other People	NUMERIC	-	57.4%	[1, 5]	3.887
R360134	R360134: I Am as Healthy as Anybody I Know	NUMERIC	-	57.36%	[1, 5]	2.934
R360135	R360135: I Expect My Health to Get Worse	NUMERIC	-	57.32%	[1, 5]	2.991
R360136	R360136: My Health Is Excellent	NUMERIC	-	57.4%	[1, 5]	3.14
SF12101	SF12101: Would You Say Your Health Is (>=WORSE)	NUMERIC	-	79.8%	[1, 5]	3.263
SF12102A	SF12102A: Health Limit Moderate Activities	NUMERIC	-	79.8%	[1, 3]	2.321
SF12102B	SF12102B: Health Limit Climb Several Flights	NUMERIC	-	79.8%	[1, 3]	2.211
SF12103A	SF12103A: Phys Health: Accomplish Less	NUMERIC	-	79.84%	[1,5]	3.501
SF12103B	SF12103B: Phys Health: Limit Kind of Work	NUMERIC	-	79.84%	[1, 5]	3.622
SF12104A	SF12104A: Emotional Problem: Accomplish Less	NUMERIC	-	79.84%	[1, 5]	3.813

VARIABILE	DESCRIZIONE	CLASSE	UM	%NA	RANGE	MEDIA
Dominio QS						
SF12104B	SF12104B: Emotional Problem: Less Careful	NUMERIC	-	79.84%	[1, 5]	3.895
SF12105	SF12105: Did Pain Interfere With Your Work	NUMERIC	-	79.84%	[1, 5]	1.96
SF12106A	SF12106A: Have You Felt Calm and Peaceful	NUMERIC	-	79.8%	[1, 5]	2.444
SF12106B	SF12106B: Did You Have a Lot of Energy	NUMERIC	-	79.8%	[1, 5]	2.819
SF12106C	SF12106C: Have Felt Downhearted/Depressed	NUMERIC	-	79.8%	[1, 5]	3.681
SF12107	SF12107: Amt Time Interfere Social Activity	NUMERIC	-	79.8%	[1, 5]	3.837
Dominio SC						
DH	Mano Dominante (DH)	CHARACTER	L/R	3.25%	-	-
Dominio OE						
SNELLEQ	Test visivo SNELLEQ: Snellen Equivalent Score, test su acuità visiva	CHARACTER	m	67.91%	-	-
Eye1.25%	SLOAN LETTER EYE CHART 1.25% , acuità visiva a basso contrasto (test NUMLCOR)	CHARACTER	-	67.95%	-	-
Eye2.5%	SLOAN LETTER EYE CHART 2.5%, acuità visiva a basso contrasto (test NUMLCOR)	CHARACTER	-	67.95%	-	-
Eye100%	SLOAN LETTER EYE CHART 100%, acuità visiva ad alto contrasto (test NUMLCOR)	CHARACTER	-	67.91%	-	-
Dominio MH						
Fenotipo	FENOTIPO: RRMS,SPMS,PPMS	CHARACTER	-	0%	-	-

Capitolo 3

Pipeline per lo sviluppo dei modelli predittivi

3.1 Pipeline riassuntiva per i modelli

Nel presente lavoro di tesi sono stati sviluppati in totale 72 modelli predittivi. Tutti i modelli hanno in comune il fatto che sono stati costruiti attorno all’algoritmo di Regressione Logistica. Inoltre, la totalità dei modelli riceve in input dati alla baseline dalle medesime 20 variabili estratte dal database MSOAC Placebo (elencate nel sottocapitolo 2.3). Sono presenti differenze tra modelli per quanto riguarda la scelta dell’orizzonte temporale di previsione, la definizione di classificazione adottata e la pipeline seguita per l’implementazione. Gli orizzonti di predizione sono stati posti a 6 mesi, 12 mesi o 18 mesi dall’inizio dello studio. Selezionato l’orizzonte temporale desiderato, i modelli seguono una tra 4 distinte definizioni di peggioramento della malattia (Δ EDSS soglia), quindi conseguono quattro diverse classificazioni dei pazienti. Una volta stabilite queste due necessità, si procede con la scelta della pipeline. Complessivamente, questa scelta può spaziare tra 6 opzioni: “Classica”, “Classica seguita da Feature Selection”, “Classica seguita da pesatura”, “Classica seguita da Feature Selection e pesi”, “Over-sampling”, “Both Under/Over-sampling”. Ciascuna pipeline viene riassunta a passi di seguito.

“Classica”

1. Separazione dei dati disponibili in Training e Test set (80% dei dati in Training set e 20% rimanente in Test set)
2. Fattorizzazione delle variabili categoriche
3. Imputazione dei valori mancanti tramite il metodo MICE
4. Normalizzazione del training set
5. Training del modello predittivo con algoritmo di Regressione Logistica sul training set imputato e normalizzato
6. Normalizzazione del Test set utilizzando informazioni del Training set
7. Applicazione del modello predittivo per predire il valore dell’outcome sul training set e sul test set
8. Calcolo curve ROC e risultati AUC per Training set e Test set

“Classica seguita da Feature Selection”

1. Separazione dei dati disponibili in Training e Test set (80% dei dati in Training set e 20% rimanente in Test set)
2. Fattorizzazione delle variabili categoriche
3. Imputazione dei valori mancanti tramite il metodo MICE
4. Normalizzazione del training set
5. Training del modello predittivo con algoritmo di Regressione Logistica sul training set imputato e normalizzato
6. Applicazione dell’algoritmo di Stepwise Feature Selection al training set (imputato e normalizzato) al fine di ricavare un modello ridotto.
7. Normalizzazione del Test set utilizzando informazioni del Training set
8. Applicazione del modello predittivo per predire il valore dell’outcome sul training set e sul test set
9. Calcolo curve ROC e risultati AUC per Training set e Test set

“Classica seguita da Pesatura”

1. Separazione dei dati disponibili in Training e Test set (80% dei dati in Training set e 20% rimanente in Test set)
2. Fattorizzazione delle variabili categoriche
3. Imputazione dei valori mancanti tramite il metodo MICE
4. Creazione di un vettore contenente i pesi
5. 5-Fold Cross Validation sul Training set (con normalizzazione Internal Training set e Internal Test set) per la scelta del peso ottimo da affidare alla classe minoritaria (Worsening)
6. Implementazione del modello predittivo con algoritmo di Regressione Logistica unito alla pesatura della classe Worsening
7. Applicazione del modello predittivo per predire il valore dell'outcome sul training set e sul test set
8. Calcolo curve ROC e risultati AUC per Training set e Test set

“Classica seguita da Feature Selection e Pesi”

1. Separazione dei dati disponibili in Training e Test set (80% dei dati in Training set e 20% rimanente in Test set)
2. Fattorizzazione delle variabili categoriche
3. Imputazione dei valori mancanti tramite il metodo MICE
4. Normalizzazione del Training set
5. Implementazione del modello predittivo con algoritmo di Regressione Logistica
6. Normalizzazione del Test set utilizzando informazioni del Training set
7. Sviluppo della Stepwise Feature Selection
8. Analisi di stabilità della Stepwise Feature Selection
9. Riduzione del Training set alle variabili stabili

10. Creazione di un vettore contenente i pesi
11. 5-Fold Cross Validation sul Training set (con normalizzazione Internal Training set e Internal Test set) per la scelta del peso ottimo da affidare alla classe minoritaria
12. Selezione delle sole variabili stabili
13. Implementazione del modello predittivo con algoritmo di Regressione Logistica unito alla pesatura della classe Worsening
14. Applicazione del modello predittivo per predire il valore dell'outcome sul training set e sul test set
15. Calcolo curve ROC e risultati AUC per Training set e Test set

“Classica con Over-sampling”

1. Separazione dei dati disponibili in Training e Test set (80% dei dati in Training set e 20% rimanente in Test set)
2. Fattorizzazione delle variabili categoriche
3. Imputazione dei valori mancanti tramite il metodo MICE
4. Over-sampling del Training set
5. Normalizzazione del training set
6. Training del modello predittivo con algoritmo di Regressione Logistica sul training set imputato e normalizzato
7. Normalizzazione del Test set utilizzando informazioni del Training set
8. Applicazione del modello predittivo per predire il valore dell'outcome sul training set e sul test set
9. Calcolo curve ROC e risultati AUC per Training set e Test set

“Classica con Both Under/Over-sampling”

1. Separazione dei dati disponibili in Training e Test set (80% dei dati in Training set e 20% rimanente in Test set)

2. Fattorizzazione delle variabili categoriche
3. Imputazione dei valori mancanti tramite il metodo MICE
4. Both Under/Over-sampling del Training set
5. Normalizzazione del training set
6. Training del modello predittivo con algoritmo di Regressione Logistica sul training set imputato e normalizzato
7. Normalizzazione del Test set utilizzando informazioni del Training set
8. Applicazione del modello predittivo per predire il valore dell'outcome sul training set e sul test set
9. Calcolo curve ROC e risultati AUC per Training set e Test set

3.2 Orizzonti temporali di previsione

Nel database MSOAC Placebo, il valore EDSS alla baseline è presente per 2385 pazienti (su un totale di 2463 individui). Man mano che ci si allontana dall'inizio dello studio, la disponibilità del dato diminuisce nel tempo. A 6 mesi dalla baseline il valore EDSS è presente per 2232 pazienti, a 12 mesi per 1674 individui, a 18 mesi per 1605 soggetti (vedi figura 2.6). In questo lavoro di tesi si è deciso di limitare a 18 mesi l'orizzonte temporale di previsione dei modelli prognostici. In tal modo, si ha a disposizione un numero adeguato di soggetti. Come facilmente intuibile, realizzando previsioni in un così ristretto arco temporale risulta difficile cogliere un peggioramento dei pazienti. Inoltre, non è confermata l'effettiva significatività clinica della previsione: un peggioramento in un arco temporale così limitato può indicare una fluttuazione casuale, soprattutto considerando l'imprevedibilità della malattia. Nonostante ciò, molti autori in letteratura ritengono rilevanti previsioni di avanzamento della malattia in 6 mesi [25] o in 24 mesi [38,39]. Altri autori spingono più avanti nel tempo le previsioni dei modelli, da 5 anni [28,36], 6 anni [27] fino a 10 anni [37]. Visionando quindi lo stato dell'arte, è possibile confermare che questo lavoro è in linea con la maggior parte degli studi. I modelli sviluppati in questa tesi spaziano le previsioni da 6 mesi a 12 e 18 mesi dall'inizio dello studio per i singoli pazienti.

3.3 Definizione di “Worsening”: le soglie

Dalla letteratura, come anticipato nel sottocapitolo 1.8, non emerge un “gold standard” di definizione del peggioramento della malattia basato sulla variazione nel tempo del valore EDSS. Diversi autori propongono la scelta per loro più adeguata. In questo lavoro, l'intento è stato quello di riportare diverse definizioni di peggioramento, prendendo in considerazione le soglie più utilizzate negli studi attualmente presenti nello stato dell'arte, per valutare il cambiamento nei risultati al variare della scelta effettuata. Di seguito sono specificate le “soglie” di classificazione (peggioramento o meno della malattia):

- Soglia 1: ΔEDSS maggiore di 0 indica un peggioramento; ΔEDSS minore di 0 indica un miglioramento/stabilità della malattia;
- Soglia 2 [28]: $\Delta\text{EDSS} \geq 1.5$ identifica un peggioramento; ΔEDSS minore di 1.5 indica che la malattia non è peggiorata;
- Soglia 3 [25]: EDSS baseline ≤ 5.5 : il paziente peggiora se ha un $\Delta\text{EDSS} \geq 1$; EDSS baseline maggiore di 6: il soggetto peggiora se ha un $\Delta\text{EDSS} \geq 0.5$;
- Soglia 4 [27]: $0.5 \leq \text{EDSS baseline} \leq 6$: il paziente peggiora se ha un $\Delta\text{EDSS} \geq 1$; EDSS baseline maggiore di 6: il soggetto peggiora se ha un $\Delta\text{EDSS} \geq 0.5$; EDSS baseline = 0: l'individuo peggiora se ha un $\Delta\text{EDSS} \geq 1.5$.

Si noti come la “Soglia 1” non garantisca con sufficiente sicurezza che per il paziente la malattia sia effettivamente avanzata. Nonostante ciò, è interessante valutare le performance dei modelli in cui è stata implementata questa soglia per poterle confrontare con le altre tre.

Selezionata la soglia di interesse, assieme all'orizzonte temporale desiderato, vengono quindi suddivisi i pazienti in due classi: “Worsening”, ovvero i pazienti che nel tempo vedono aggravarsi la malattia, e “Non Worsening”, inclusiva dei pazienti che rimangono stabili o migliorano nel tempo.

3.4 Split in Training set e Test set

In questo lavoro è stata scelta una suddivisione 80/20 del dataset: l'80% dei pazienti viene affidato al così detto “Training” set, ovvero la porzione di dati con la quale il modello viene allenato; il 20% rimanente diventa invece il “Test” set, parte dei dati che viene utilizzata unicamente alla fine del lavoro per la validazione dei risultati

ottenuti. Questa separazione è necessaria, poiché se il modello viene allenato su tutti i dati a disposizione non si ha più la possibilità di validarlo con dati che il modello stesso non ha mai visto. Importante ricordare che durante questo split dei dati sono state mantenute in Training set e Test set le medesime percentuali di pazienti di ciascuna classe, in modo tale da non creare errori o bias nella distribuzione degli individui tra i due set. Al variare dell’orizzonte temporale selezionato, cambia il numero di soggetti a disposizione per ciascun modello. Nella tabella 3.1 viene riportato il numero di pazienti nel training set e nel test set al variare di dataframe alla baseline.

Tabella 3.1: Soggetti spartiti tra Training set e Test set al variare dell’orizzonte temporale di previsione scelto

	DF completo baseline	Training set	Test set
6 Mesi	2232	1786	446
12 Mesi	1674	1339	335
18 Mesi	1605	1284	321

3.5 Fattorizzazione delle variabili categoriche

Il tipo di ciascuna feature è espresso in tabella 2.2. Per sviluppare i modelli computazionali è stato necessario fattorizzare alcune delle covariate presenti nel dataset ottenuto dalla fase di pre-processing. La fattorizzazione è un processo obbligatorio per alcune variabili e indispensabile per la successiva imputazione dei dati e per l’implementazione dell’algoritmo di Regressione Logistica. Come intuibile, le variabili categoriche “SEX”, “FENOTIPO” e “DH” sono state fattorizzate tramite il metodo tradizionale: viene associato un livello casuale per ciascun tipo di realizzazione della variabile. Nel caso della variabile “SEX” si ottengono quindi due livelli ($M = 0$, $F = 1$). Ugualmente anche la variabile “DH” è separata in due possibili livelli ($L = 1$, $R = 0$). Nel caso invece della variabile “FENOTIPO” viene associato un livello ad ogni fenotipo, di conseguenza si ricava una variabile a tre livelli ($RRMS = 0$, $SPMS = 0.5$ e $PPMS = 1$). La variabile “AGE” e le feature estratte dal dominio “FT” sono state mantenute di tipo “Numeric”. E’ stata effettuata invece una fattorizzazione *ordinata* per la variabile “P1Y” e le feature ricavate dal dominio “QS” (EDSS e le componenti del questionario KFSS). In questo caso le variabili sono fattorizzate a livelli che mantengono l’importanza dell’ordine tra loro. Medesimo discorso vale per i questionari: il risultato della fattorizzazione è ordinale (0 indica la condizione

migliore per il soggetto; aumentando il modulo del livello peggiora la condizione del paziente). La variabile P1Y è stata suddivisa in 6 livelli (“0”, “1”, “2”, “3”, “4” e “ maggiore di 4 ”). Le variabili “KFSS” sono suddivise tra i 4 e i 7 livelli, in base ai risultati (se vanno da 0-4, 0-5 o 0-7). Il parametro EDSS (range 0-6.5), infine, è stato fattorizzato in 13 livelli.

3.6 Imputazione dei dati mancanti: MICE

I dati mancanti (Not Availables, NAs) sono un problema ricorrente in qualunque ambito statistico. Un dato può essere assente per svariati motivi, per esempio a causa di una dimenticanza da parte del personale o irreperibilità del paziente per una visita. Per poter sfruttare al massimo il potenziale dei dati a disposizione, esistono ad oggi numerose tecniche di “imputazione” dei dati. Questi metodi consentono di attribuire un valore laddove il dato non sia presente. La tecnica più semplice in assoluto è quella di imputazione dei “missing values” di una variabile con la media dei dati preesistenti della stessa. Questo procedimento però non considera la correlazione tra le variabili del dataset. In questo lavoro di tesi si è deciso di adottare un metodo all’avanguardia, più complesso, che tiene conto della correlazione tra feature, realizzando per cui imputazioni affidabili e plausibili: MICE (Multivariate Imputation by Chained Equations). Il metodo, partendo da un’imputazione semplice (es. utilizzando i valori medi delle variabili) procede poi con imputazioni multiple (in maniera ricorsiva) che conducono ad un affinamento dei valori conferiti ai dati mancanti [57].

L’imputazione tramite MICE è stata implementata nel software RStudio (versione 1.4.1717) utilizzando la funzione apposita “mice” presente nell’omonima libreria. Grazie alla precedente feature selection fondata sul numero di NAs di ogni variabile, l’imputazione è stata realizzata unicamente per feature con pochi valori mancanti. In tabella 3.2 vengono riprese le percentuali di NAs per ogni feature del dataset alla baseline (senza le variabili con 0% di valori mancanti, le quali non necessitano di imputazione). La variabile con più valori da imputare è la feature P1Y: circa 30% di NAs. Per questa cavariata è stato mostrato l’andamento discreto dei livelli pre imputazione e post imputazione (figura 3.1), in cui si può apprezzare come il metodo MICE riesca a creare imputazioni ben distribuite tra i livelli della variabile. Tramite MICE, si è giunti per cui infine alla creazione di Training set e Test set completi dei dati, quindi utilizzabili rispettivamente come input e per la validazione di modelli predittivi.

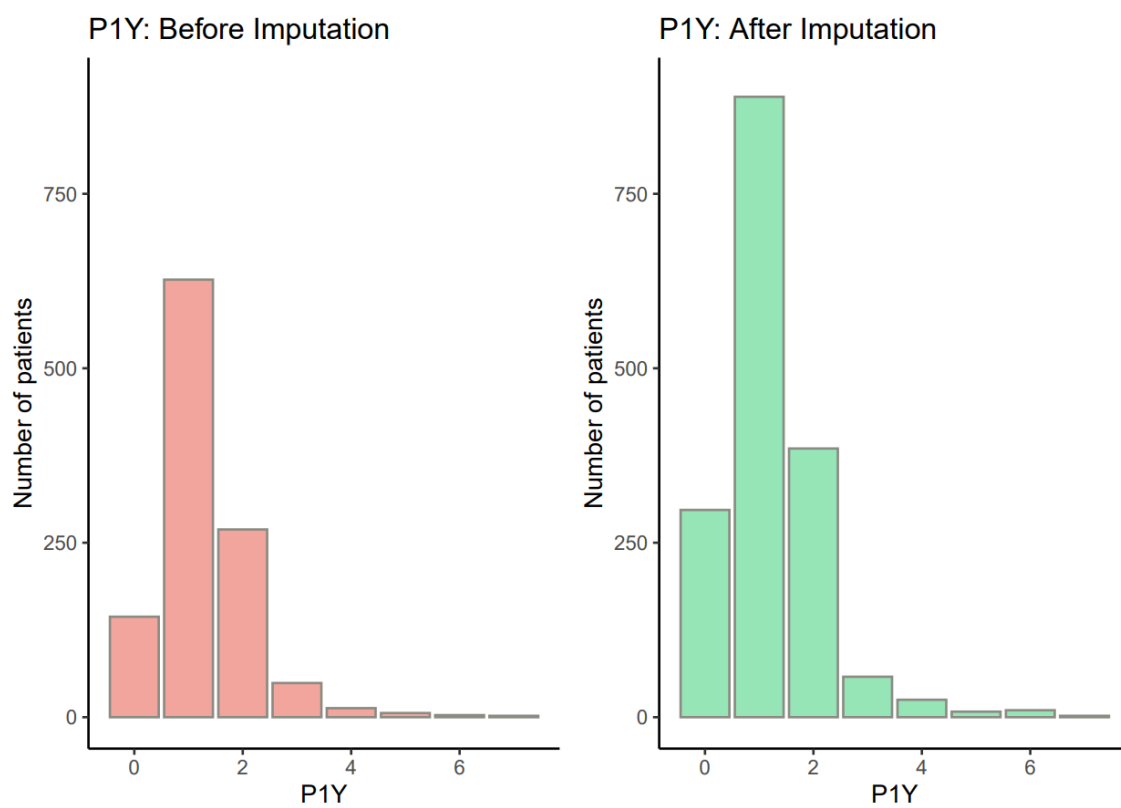


Figura 3.1: Pre e post imputazione della variabile con il maggior numero di valori mancanti: P1Y. La percentuale di NAs nel data frame per questa feature è pari al 27.75%. P1Y è l'unica variabile con una percentuale così elevata di valori mancanti, tutte le altre si assestano al di sotto del 17%. Modelli di riferimento: 12 mesi.

Tabella 3.2: Percentuali di NAs per ciascuna feature imputata del dataset alla baseline

VARIABILE	DOMINIO	PERCENTUALE NAs
AGE	Demographics	3.37%
P1Y	Foundings About	27.75%
T25FW_1	Functional Tests	0.81%
T25FW_2	Functional Tests	0.85%
T25FW_1	Functional Tests	0.81%
PASAT_tc_3s	Functional Tests	3.41%
NHPT01_DH1	Functional Tests	3.41%
NHPT01_DH2	Functional Tests	3.45%
NHPT01_NDH1	Functional Tests	3.77%
NHPT01_NDH2	Functional Tests	3.85%
EDSS	Questionnaires	3.29%
KFSS_01	Questionnaires	15.62%
KFSS_02	Questionnaires	16.35%
KFSS_03	Questionnaires	15.62%
KFSS_04	Questionnaires	15.62%
KFSS_05	Questionnaires	15.62%
KFSS_06	Questionnaires	15.62%
KFSS_07	Questionnaires	15.62%
DH	Subject Characteristics	3.25%

3.7 Normalizzazione delle variabili

Ogni variabile ha la propria unità di misura e di conseguenza uno specifico modulo per la varianza. Affinché la grandezza dei dati non influisca in alcun modo sui risultati, è indispensabile normalizzare tutte le covariate. Nel presente lavoro, per tale motivo, è stata effettuata una normalizzazione delle variabili. Inoltre, in tal modo è possibile eseguire confronti tra i coefficienti associati dal modello alle varie covariate. La normalizzazione del Training set è stata eseguita con la formula sottostante, che consente di racchiudere i dati di ciascuna variabile nell'intervallo 0-1: a ciascun dato viene sottratto il massimo della variabile di appartenenza e il tutto viene poi diviso per la differenza tra massimo e minimo della variabile.

$$\frac{x - \min}{\max - \min}$$

La normalizzazione del Test set è stata eseguita secondo la medesima formula ma

utilizzando massimi e minimi delle variabili salvati dal Training set. L'intervallo di variabilità delle feature può quindi eccedere il massimo di 1, ma questo non conduce ad alcun problema per l'implementazione dei modelli.

3.8 Regressione Logistica

L'algoritmo utilizzato per sviluppare i modelli predittivi è quello di Regressione Logistica (Logistic Regression, LR). Questo algoritmo consente una classificazione binaria: invece di agire come un classico algoritmo di regressione (per es. lineare) che modella direttamente l'output dalle variabili indipendenti, la Regressione Logistica stima la probabilità che la variabile dipendente appartenga ad una delle due classi impostate. Questa tecnica statistica risulta per cui ideale per lo scopo della tesi: classificazione pazienti MS nelle classi "Worsening" e "Non Worsening". In questa tesi la Regressione Logistica è stata implementata con la funzione "glm" (Generalized Linear Models) di R ponendo il parametro di input "family = Binomial".

3.9 Stepwise Feature Selection

Con il termine "Feature Selection" si intende una scrematura del numero di variabili, la quale consente di mantenere unicamente quelle con il maggior potere predittivo per il modello. I vantaggi di questa metodologia sono numerosi: il modello diviene più snello, semplice da interpretare e pratico da implementare. Inoltre, se il modello sfrutta covariate irrilevanti dal punto di vista predittivo, oltre ad essere inutilmente più complesso, rischia di avere performance inferiori. Svitati metodi di feature selection sono stati sviluppati al giorno d'oggi; in questa tesi si è optato per l'utilizzo della tecnica "Stepwise", la quale esiste in 3 forme:

- Forward: partendo dal modello "nullo", ovvero senza alcuna variabile, il metodo forward aggiunge man mano le covariate, valutando nel mentre la capacità predittiva del modello; ad ogni step viene aggiunta al modello la variabile che risulta più significativa per la predizione.
- Backward: l'algoritmo ha inizio dal modello "completo", ovvero con tutte le covariate; iterativamente rimuove le variabili che non migliorano le performance del modello;
- Both: il metodo unisce le precedenti tecniche forward e backward.

Testando tutte e tre le possibili forme del metodo si è giunti alla decisione di utilizzare la tecnica “Backward”, la quale ha condotto a prestazioni migliori da parte di tutti i modelli. L’analisi di stabilità delle variabili, presente in una delle pipeline implementate, viene realizzata tramite una 5-fold cross validation eseguita per la Stepwise Feature Selection. Sono state mantenute le feature presenti in almeno 3 delle 5 fold della cross validazione.

3.10 Sbilanciamento tra classi

Un importante problema da tenere in considerazione, quando si lavora in ambito di classificazione, è quello dello sbilanciamento tra le classi. Nella totalità degli articoli dello stato dell’arte di modelli predittivi per prevedere il peggioramento di pazienti affetti da Sclerosi Multipla, le classi sono fortemente sbilanciate tra loro. La maggior parte dei pazienti non peggiora durante l’arco di tempo indagato, il che porta ad un numero molto più elevato di pazienti che rimangono stabili o migliorano nel tempo. Ciò accade soprattutto quando le previsioni si spingono non oltre i 2-3 anni, tempo troppo breve per poter cogliere significativi peggioramenti nei pazienti. Lo sbilanciamento tra le classi conduce a un bias nelle predizioni che il modello tenta di realizzare. Quando il modello si allena sul Training set, se vede molti più casi di “Non Worsening” tra i soggetti, quando poi viene testato sul Test set tende ad associare più individui alla classe che ha maggiormente incontrato nel Training set. Per risolvere questo considerevole ostacolo, al giorno d’oggi le tecniche più utilizzate sono:

- Over-Sampling: la classe con il numero inferiore di osservazioni viene portata alla dimensione dell’altra classe, tramite la replica dei record presenti nella classe “minore” [58,59];
- Under-Sampling: la classe “maggiore” viene ridotta in dimensione fino a parreggiare la classe con il numero minore di osservazioni. Questa tecnica può essere applicata unicamente quando il numero di dati a disposizione è sufficientemente elevato (si perdono definitivamente i dati rimossi) [58,59];
- Both Under/Over-Sampling: Mix tra le due tecniche precedenti, si riduce la dimensione della classe maggiore e si aumenta quella della classe inferiore; [58,59]
- Weighting: alle due classi viene affidato un peso differente. Questo implica che l’algoritmo (in questo caso la Regressione Logistica) andrà ad affidare una

maggior importanza alle osservazioni che appartengono alla classe con il peso maggiore. In questo metodo viene quindi dato un peso più grande alla classe minoritaria, in modo da cercare il bilanciamento con la classe più grande.

Nella presente tesi sono stati testati tutti questi metodi: tramite le tecniche “Over-sampling” e “Both Under/Over-sampling” le classi sono state portate ad un rapporto paritario. La tecnica di “Under-Sampling” è stata esclusa per evitare di diminuire eccessivamente il numero di dati a disposizione. Infine, il metodo “Weighting” è stato implementato utilizzando una 5 Fold Cross Validazione per la scelta del peso da affidare alla classe “Worsening”. L’algoritmo costruito può essere schematizzato in tal modo:

1. Creazione di un vettore contenente i pesi da affidare alla classe minoritaria; i pesi introdotti per la successiva cross validazione sono: 1, 1.2, 1.5, 2, 3, 5, 8, 10, 11, 13, 16, 20, 25, 30, 40, 50, 75, 100, 125, 150.
2. Suddivisione del Training Set in 5 parti (ogni parte è costituita dal Training Set Interno, 4/5 del Training Totale, e il Test Set Interno, il rimanente 1/5 del Totale)
3. 5-Cross Validazione con normalizzazione interna del Training Interno e del Test Interno e sviluppo del modello di Regressione Logistica con pesatura
4. Calcolo della Media dei risultati AUC ottenuti tra le 5 folds al variare del peso selezionato
5. Scelta del peso corrispondente al valore AUC medio più elevato

In seguito sono stati effettuati dei confronti della capacità, da parte di ciascun metodo, di migliorare le performance dei modelli.

Capitolo 4

Risultati

4.1 Modelli per predizioni a 6 mesi

I modelli che realizzano la classificazione a 6 mesi dalla baseline hanno a disposizione in input dati di 2232 pazienti del database MSOAC Placebo. Il numero di pazienti per ciascuna classe al variare della soglia selezionata è mostrato in tabella 4.1. In termini di “Area Under The ROC” (AUC), il risultato migliore si è trovato tramite il modello con soglia 2 (peggioramento identificato tramite un aumento pari o maggiore a 1.5 sulla scala EDSS) che segue la pipeline di “Over-sampling” (AUC = 0.74). Per i 24 modelli distinti che effettuano predizioni a 6 mesi dalla baseline, i risultati AUC del Training set e Test set sono riportati nelle tabelle 4.2 (Soglia 1), 4.3 (Soglia 2), 4.4 (Soglia 3) e 4.5 (Soglia 4). In seguito ai risultati viene mostrato il grafico della curva ROC (Figura 4.1) per il modello che ha conseguito la migliore performance tra i 24 sviluppati. Assieme a questa figura viene riportata anche la lista dei migliori predittori per questo modello con annessi p-values (Tabella 4.6). Come si può notare, tra le variabili demografiche risulta di grande utilità l’età dei soggetti. Per quanto riguarda le feature cliniche, invece, sono particolarmente rilevanti la variabile P1Y, il valore EDSS e i test funzionali T25FW, PASAT e NHPT.

Tabella 4.1: Sbilanciamento classi al variare della soglia impostata (numero soggetti per classe). Modelli di predizioni a 6 mesi dalla baseline

	SOGLIA 1	SOGLIA 2	SOGLIA 3	SOGLIA 4
WORSENING	506	103	277	222
NON WORSENING	1726	2129	1955	2010
Total Unbalancing	22.7/77.3%	4.6/95.4%	12.4/87.6%	10/90%

Tabella 4.2: **Soglia 1**: risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.7013	0.6509
Classica + FS	0.6788	0.644
Classica + weight (8)	0.7029	0.6503
Classica + FS + weight (125)	0.6725	0.6396
Over-sampling	0.7075	0.6162
Both U/O-sampling	0.7425	0.6338

Tabella 4.3: **Soglia 2**: risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.8232	0.737
Classica + FS	0.7492	0.7081
Classica + weight (50)	0.8423	0.7364
Classica + FS + weight (2)	0.7492	0.7072
Over-sampling	0.8457	0.7404
Both U/O-sampling	0.8504	0.7284

Tabella 4.4: **Soglia 3**: risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.76	0.6083
Classica + FS	0.7281	0.5997
Classica + weight (8)	0.7661	0.6149
Classica + FS + weight (5)	0.7221	0.5907
Over-sampling	0.7697	0.6052
Both U/O-sampling	0.7799	0.5912

Tabella 4.5: **Soglia 4**: risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.7298	0.6275
Classica + FS	0.6708	0.6523
Classica + weight (125)	0.7317	0.637
Classica + FS + weight (25)	0.6712	0.6512
Over-sampling	0.7516	0.6325
Both U/O-sampling	0.7518	0.609

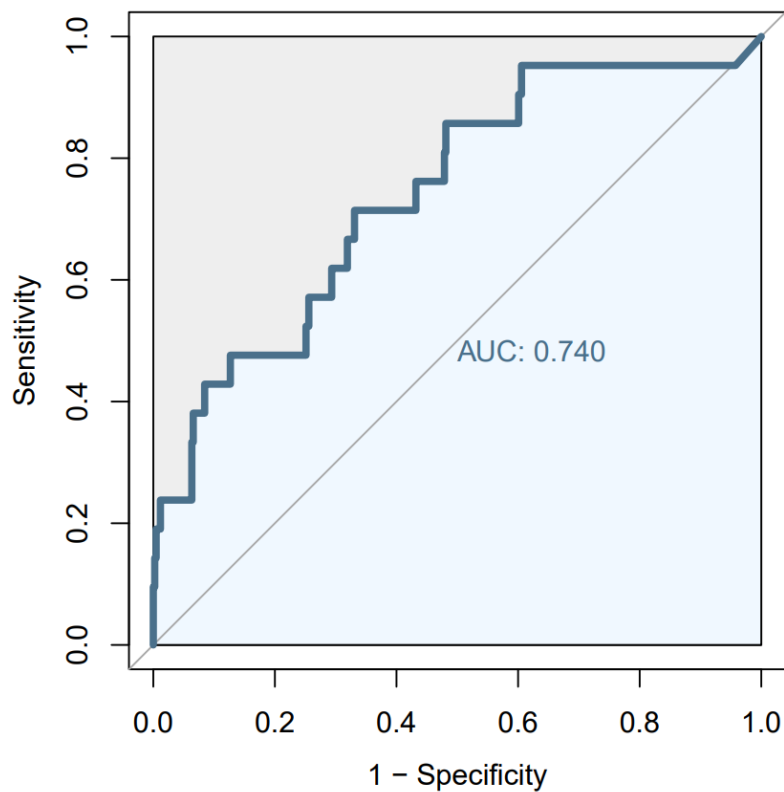


Figura 4.1: Curva ROC per il modello a 6 mesi con pipeline “Over-sampling” in cui la soglia utilizzata è la numero 2. Il risultato AUC è pari a 0.7401 con un intervallo di confidenza tra 0.626 e 0.855

Tabella 4.6: Migliori predittori per il modello con soglia 2 e pipeline “Classica”

Variabile	Estimate	p-value
AGE	2.41799	4.95e-15
P1Y	3.04562	1.2×10^{-16}
T25FW_1	-36.97893	7.97e-05
T25FW_2	45.84491	9.14e-06
PASAT	-2.34496	1.2×10^{-16}
NHPT_DH1	24.66315	4.72e-06
EDSS	-2.32802	8.70e-12
Fenotipo	0.89214	0.000287

4.2 Modelli per predizioni a 12 mesi

Cercando predizioni a 12 mesi dall’inizio dello studio, lo sbilanciamento tra classi migliora lievemente. Si può notare il maggior equilibrio tramite i dati riportati in tabella 4.7. Per questi modelli (24 in totale, 6 pipeline differenti per ciascuna soglia scelta) il numero di pazienti disponibili è pari a 1674. La migliore performance è realizzata dal modello che utilizza la Soglia 4 (vedi sezione 3.3) e segue la pipeline “Classica seguita da Feature Selection con Pesi”. Il peso affidato alla classe Worsening, selezionato tramite la 5-fold cross validation, è pari a 5: un’osservazione di un paziente che peggiora viene considerata cinque volte più rilevante rispetto ad un record di un soggetto della classe Non Worsening. Per questo modello si raggiungono prestazioni al valore AUC di 0.7121. Come eseguito per la sezione 4.1, anche per i modelli “a 12 mesi” vengono riportate 4 tabelle che esprimono i risultati AUC di Training set e Test set al variare della Soglia selezionata: tabella 4.8 per i risultati con la Soglia 1, tabella 4.9 relativa alla Soglia 2, tabella 4.10 alla Soglia 3 ed infine 4.11 per la quarta Soglia. A fine del corrente sottocapitolo viene esposto il grafico ROC per il modello migliore (in termini di AUC, figura 4.2) e i relativi best predictors in tabella 4.12. Questi predittori (AGE, T25FW_1, NHPT_NDH2, EDSS, KFSS_05 e l’intercetta del modello), data la pipeline seguita dal modello, sono i migliori in termini di p-value associato tra quelli ricavati dopo aver eseguito l’analisi di stabilità della Stepwise Feature Selection.

Tabella 4.7: Sbilanciamento classi al variare della soglia impostata (numero soggetti per classe). Modelli di predizioni a 12 mesi dalla baseline

	SOGLIA 1	SOGLIA 2	SOGLIA 3	SOGLIA 4
WORSENING	523	120	322	250
NON WORSENING	1151	1554	1352	1424
Total Unbalancing	31.2/68.8%	7.2/92.8%	19.2/80.8%	15/85%

Tabella 4.8: **Soglia 1:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.6849	0.5665
Classica + FS	0.6609	0.5826
Classica + weight (1.2)	0.6851	0.5658
Classica + FS + weight (13)	0.6516	0.5824
Over-sampling	0.707	0.5732
Both U/O-sampling	0.726	0.5561

Tabella 4.9: **Soglia 2:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.8386	0.6058
Classica + FS	0.8173	0.5518
Classica + weight (1)	0.8386	0.6058
Classica + FS + weight (1)	0.7925	0.5965
Over-sampling	0.8582	0.6034
Both U/O-sampling	0.8902	0.5982

Tabella 4.10: **Soglia 3:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.712	0.62
Classica + FS	0.6956	0.6578
Classica + weight (1)	0.712	0.62
Classica + FS + weight (150)	0.6965	0.6616
Over-sampling	0.7241	0.6055
Both U/O-sampling	0.721	0.627

Tabella 4.11: **Soglia 4:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.7154	0.6751
Classica + FS	0.6976	0.7075
Classica + weight (40)	0.7181	0.6454
Classica + FS + weight (5)	0.6997	0.7121
Over-sampling	0.7176	0.689
Both U/O-sampling	0.7741	0.6834

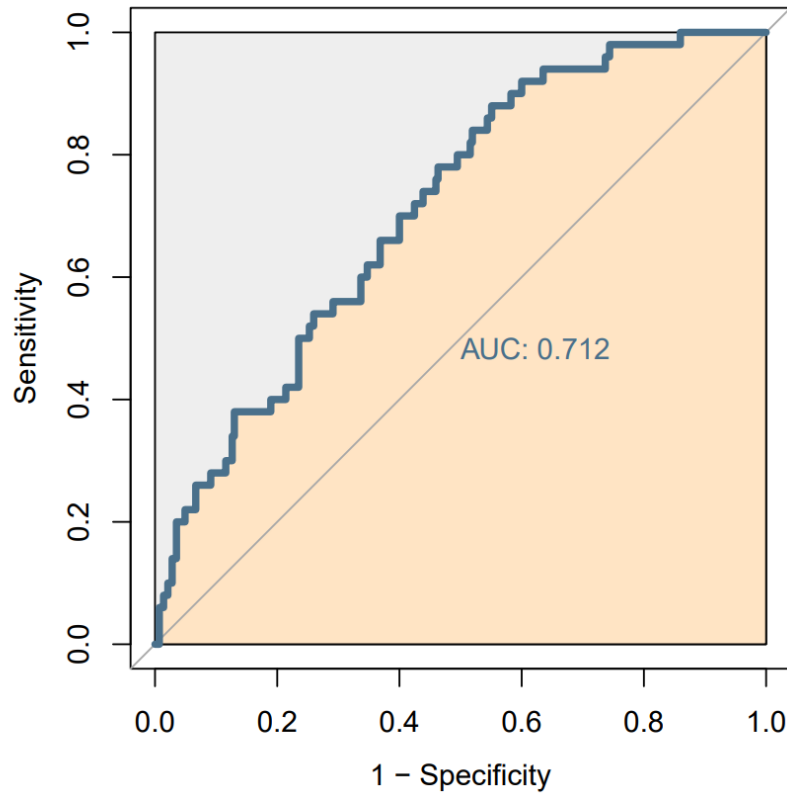


Figura 4.2: Curva ROC per il modello a 12 mesi con pipeline “Classica con feature selection e pesatura” in cui la soglia utilizzata è la numero 4. Il peso affidato alla classe Worsening è “5.” Il risultato AUC è pari a 0.712 con un intervallo di confidenza tra 0.641 e 0.783

Tabella 4.12: Migliori predittori per il modello con soglia 4 e pipeline “Classica con Feature Selection e pesatura”

Variabile	Estimate	p-value
Intercept	-1.1110	1.73e-06
AGE	0.9622	0.000382
T25FW_1	6.0852	8.12e-06
NHPT_NDH2	3.1306	0.004381
EDSS	-2.0192	1.25e-12
KFSS_05	0.7287	0.001267

4.3 Modelli per predizioni a 18 mesi

Per concludere i risultati ottenuti, in questa sezione vengono mostrate le performance dei modelli che effettuano la classificazione in base al risultato Δ EDSS a 18 mesi dall'inizio dello studio. Questi 24 modelli risultano i più completi: lo sbilanciamento tra classi diminuisce (nonostante sia ugualmente preponderante) poiché l'orizzonte temporale si estende. Inoltre, predizioni a quasi due anni dalla baseline risultano effettivamente più utili, dal punto di vista clinico, delle previsioni precedentemente realizzate (6 mesi e 12 mesi). Il numero di pazienti presenti per lo sviluppo dei modelli "a 18 mesi" sono in totale 1605. Come nei precedenti sottocapitoli, vengono riportati i risultati tramite tabelle: la tabella 4.14 mostra i risultati per i modelli che seguono la sogliatura 1, la tabella 4.15 descrive le performance dei modelli che utilizzano la Soglia 2, la tabella 4.16 espone i risultati trovati con la sogliatura 3 ed infine in tabella 4.17 sono sintetizzate le prestazioni dei modelli ai quali è applicata la Soglia 4. Il miglior modello, sempre in termini di risultato AUC sul Test set è il modello soglia con la soglia 2 in cui viene applicata la tecnica di Both Under/Over sampling. Tramite questa tecnica di bilanciamento si ottengono due classi con eguale dimensione.

Il risultato AUC sul Test set del modello appena citato è pari a 0.7101. Il grafico che descrive la curva ROC per tale modello è riportato in figura 4.3. Gli annessi "Best Predictors" con relativi "estimates" e "p-values" sono esposti in tabella 4.18. Si può notare che per il modello risultano particolarmente rilevanti lo score EDSS e il fenotipo dei pazienti.

Tabella 4.13: Sbilanciamento classi al variare della soglia impostata (numero soggetti per classe). Modelli di predizioni a 18 mesi dalla baseline

	SOGLIA 1	SOGLIA 2	SOGLIA 3	SOGLIA 4
WORSENING	571	164	390	320
NON WORSENING	1034	1441	1215	1285
Total Unbalancing	35.6/64.4%	10.2/89.8%	24.3/75.7%	20/80%

Tabella 4.14: **Soglia 1:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.6872	0.5985
Classica + FS	0.6618	0.617
Classica + weight (10)	0.6869	0.5875
Classica + FS + weight (1.5)	0.6639	0.5976
Over-sampling	0.6992	0.5741
Both U/O-sampling	0.7305	0.5881

Tabella 4.15: **Soglia 2:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.7801	0.7064
Classica + FS	0.7391	0.7002
Classica + weight (75)	0.7903	0.707
Classica + FS + weight (1)	0.7379	0.6602
Over-sampling	0.7904	0.6969
Both U/O-sampling	0.8258	0.7101

Tabella 4.16: **Soglia 3:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.6761	0.6618
Classica + FS	0.6442	0.6363
Classica + weight (11)	0.6812	0.6446
Classica + FS + weight (1.5)	0.6421	0.6445
Over-sampling	0.691	0.6496
Both U/O-sampling	0.7382	0.6177

Tabella 4.17: **Soglia 4:** risultati AUC Training set e Test set al variare della pipeline seguita per l'implementazione del modello

Pipeline	AUC Training	AUC Test
Classica	0.703	0.5945
Classica + FS	0.68	0.5913
Classica + weight (3)	0.7082	0.6113
Classica + FS + weight (1.2)	0.676	0.5789
Over-sampling	0.7331	0.605
Both U/O-sampling	0.7453	0.5937

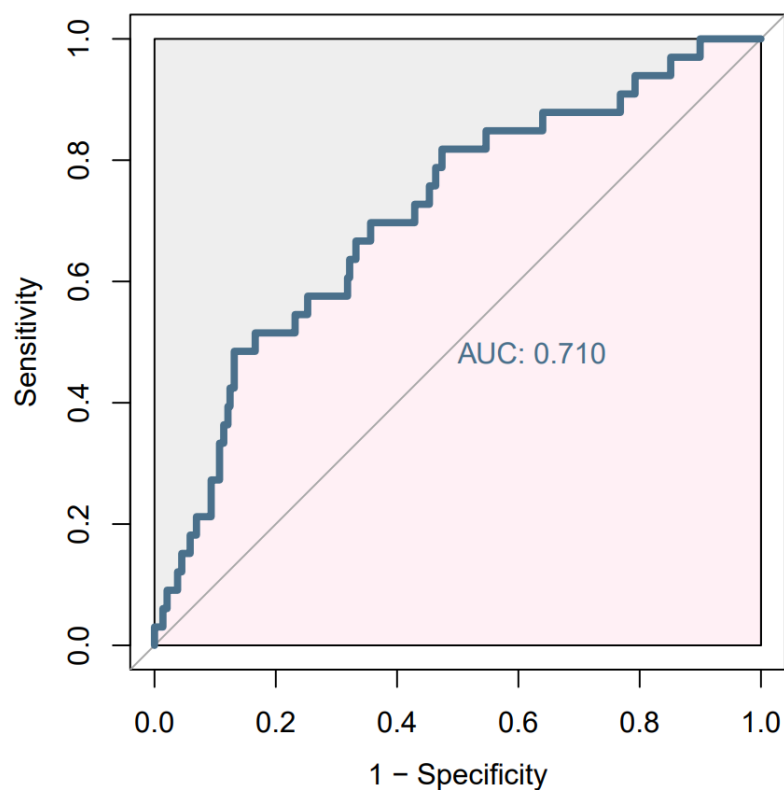


Figura 4.3: Curva ROC per il modello a 12 mesi con pipeline “Both Under e Over sampling” in cui la soglia utilizzata è la numero 2. Il risultato AUC è pari a 0.7101 con un intervallo di confidenza tra 0.615 e 0.805

Tabella 4.18: Migliori predittori per il modello con soglia 2 e pipeline “Classica”

Variabile	Estimate	p-value
P1Y	-0.702145	0.026148
NHPT_DH1	8.659102	0.012367
EDSS	-2.606202	7.33e-11
KFSS_01	-2.732400	0.000386
KFSS_03	-1.703099	0.056166
KFSS_05	1.549775	0.006603
DH	0.521054	0.030788
Fenotipo	-1.219207	2.11e-05

Capitolo 5

Conclusioni e sviluppi futuri

In questo lavoro di tesi sono state applicate tecniche di Machine Learning con l'obiettivo di classificare pazienti affetti da Sclerosi Multipla in base al riscontro di un peggioramento o meno della malattia in un prestabilito arco temporale dalla baseline (qui intesa come l'inizio dello studio). I modelli predittivi implementati per eseguire la classificazione sono costruiti tramite dati estratti dal database MSOAC Placebo, comprendente 2465 pazienti affetti da Sclerosi Multipla. Questo database rispecchia in maniera eccellente una classica popolazione di individui colpiti da Sclerosi Multipla. Dal punto di vista demografico, nel database è rispettato il rapporto 2:1 tra sessi riportato in letteratura [2,5]: 1658 femmine e 807 maschi. Inoltre, l'età dei pazienti è ben distribuita tra 18 e 72 anni: un ampio range che riflette un'ottima omogeneità dell'età nel campione di studio. Per quanto riguarda la suddivisione dei fenotipi, nel database il 64% dei soggetti è affetto dalla forma RRMS (in letteratura confermato come fenotipo più frequente [11]), il 23% degli individui è colpito dal fenotipo SPMS e il 13% dei pazienti ha la forma PPMS della patologia (in letteratura viene riportata una frequenza del 15% di questa forma nella popolazione [5]). Nel database MSOAC Placebo sono salvati risultati di test funzionali considerati tra i più significativi per cogliere gli effetti della malattia: il test T25FW per valutare la camminata dei soggetti [45], il test NHPT per analizzare la destrezza manuale [49,50] e il PASAT-3s per esaminare le prestazioni dei pazienti a livello cognitivo [48]. Infine, per quanto concerne lo score EDSS, l'intervallo di variabilità risulta tra 0 e 6.5 all'inizio dello studio (range di variabilità in cui il test T25FW è considerato significativo a livello clinico [45]). Per questi motivi, il presente lavoro conduce ad una rilevante validità dei risultati successivamente ottenuti.

Come intuibile dal nome del database stesso (MSOAC "Placebo"), tutti i pazienti appartengono a rami placebo degli studi di provenienza. Di conseguenza, in questo lavoro non sono valutati effetti di terapie o trattamenti somministrabili ai pazienti.

Nel database MSOAC Placebo non sono presenti informazioni di carattere para-clinico (per esempio dati MRI), spesso utilizzate negli studi nello stato dell'arte [25,27,28,36,37,38].

Durante la fase di pre-processing, dal database sono state ricavate 20 feature alla baseline, fornite come input ai modelli predittivi: AGE, SEX, P1Y (numero di ricadute nell'anno precedente alla baseline), 2 variabili per il test T25FW, quattro per il test NHPT, PASAT, EDSS, 7 variabili per il questionario KFSS, Fenotipo e DH (mano dominante).

In questo lavoro sono stati indagati 3 differenti orizzonti temporali di previsione del peggioramento della malattia: 6 mesi dalla baseline (2232 pazienti a disposizione), 12 mesi dalla baseline (1674 pazienti a disposizione) e 18 mesi dalla baseline (1605 pazienti a disposizione). Il numero di pazienti, a prescindere dall'arco temporale di previsione, è superiore rispetto alla maggior parte degli studi nello stato dell'arte [25,27,28,36,37,38]. La cardinalità è simile invece al lavoro di Zhao et al. [36], nel quale gli autori impiegano dati di 1693 pazienti.

Sono stati implementati 24 modelli predittivi per ciascun orizzonte temporale di previsione, per un totale di 72 modelli. Ciascuno di essi effettua una classificazione dei pazienti nelle classi "Worsening" (classe a cui vengono assegnati pazienti che peggiorano nel tempo) e "Non Worsening" (classe inclusiva di pazienti che migliorano o sono stabili nel tempo) tramite l'algoritmo di Regressione Logistica.

Un'importante novità introdotta in questo lavoro di tesi è l'utilizzo di distinte definizioni di peggioramento della malattia, tutte fondate su una variazione temporale del valore EDSS. In tal modo, si sono potute confrontare le performance dei modelli al variare della definizione scelta. Si osserva che impostata una definizione di peggioramento, i risultati ottenuti, al variare della pipeline seguita per implementare i modelli, sono affini: questo suggerisce che la scelta della definizione di peggioramento è centrale per le prestazioni conclusive dei modelli. Considerando i risultati conseguiti, per collimare la necessità di un "gold standard" per la comunità scientifica, in questo lavoro si consiglia l'utilizzo della definizione di peggioramento proposta da Zhao et al. [36]: un paziente "peggiora" quando è presente un aumento del valore EDSS di almeno 1,5 (nello studio individuato in un arco temporale di 5 anni).

Come spiegato nella sezione 3.2, gli orizzonti temporali scelti sono in linea con la letteratura, in cui spesso si indagano archi temporali da 6 mesi [25] fino a 2 anni [38,39]. In un così ristretto intervallo di tempo (soprattutto a 6 mesi dalla baseline) è però improbabile cogliere il peggioramento dei pazienti. Ciò comporta quindi un forte sbilanciamento delle classi. Nel presente lavoro si spazia da un rapporto Worsening/Non Worsening all'incirca di 5:95 a rapporti all'incirca di 35:65. In ag-

giunta, più breve è l'orizzonte temporale di previsione, minore è l'utilità clinica della predizione, poichè gli attuali trattamenti sono comunemente a lungo termine. Per sovvenire al disequilibrio tra le due classi, in questa tesi sono state implementate 3 tecniche di bilanciamento: Over-sampling, Both Under/Over-sampling e pesatura della classe di dimensione inferiore. In generale, i risultati migliori sono ottenuti per modelli in cui una tra queste tecniche viene impiegata. Questo lavoro rimarca come un'attenta gestione delle situazioni in cui le classi sono fortemente sbilanciate sia fondamentale per consentire il raggiungimento di buone performance di predizione. In questo studio, i risultati delle performance dei modelli sono calcolati mediante il valore AUC (modalità seguita da molti autori nello stato dell'arte [25,27,28,37,39]). Il modello migliore per predizioni a 6 mesi dalla baseline ha raggiunto un valore $AUC = 0.74$. I modelli migliori per predizioni a 12 e 18 mesi dalla baseline riportano un valore $AUC = 0.71$. Questi risultati appaiono notevolmente inferiori rispetto ad alcuni studi [27,28,38]. In questo lavoro si ipotizza che la mancanza di variabili paracliniche sia alquanto influente; variabili di imaging sono difatti spesso considerate fondamentali per garantire ottime prestazioni da parte dei modelli. Tuttavia i modelli sviluppati in questa tesi risultano avere delle performance superiori rispetto ai modelli degli studi [25,36] che utilizzano l'algoritmo di Regressione Logistica per effettuare predizioni entro rispettivamente 5 anni e 6 mesi dalla baseline. Si ipotizza che questo sia dovuto alla cardinalità dei dati, in questo lavoro maggiore rispetto ai lavori a confronto.

Una lista dei migliori predittori è stata stilata per i 3 modelli migliori (uno per ciascun arco temporale di previsione). Le variabili più rilevanti per le predizioni dei modelli sono risultate lo score EDSS e il test sulla mobilità degli arti superiori NHPT, già riscontrate in altri studi come feature essenziali [25,27,28,36,37,38].

Nonostante le performance dei modelli implementati non possano garantirne un utilizzo in ambito clinico, questo lavoro di tesi ha rimarcato la grande potenzialità dell'utilizzo di modelli predittivi per prevedere l'evoluzione della Sclerosi Multipla. Possibili sviluppi futuri a questo lavoro sono l'utilizzo di diversi algoritmi di classificazione, rispetto al qui utilizzato algoritmo di Regressione Logistica (ad esempio Support Vector Machine o Random Forest); l'impiego di un numero maggiore di dati (per valutare se questo possa garantire ottime prestazioni dei modelli anche in assenza di variabili paracliniche); l'estensione dell'orizzonte temporale di previsione, in modo tale da condurre ad una maggiore significatività clinica delle previsioni e al tempo stesso garantire un adeguato bilanciamento tra le classi; ed infine l'incisione di variabili paracliniche tra i dati forniti in input ai modelli (per esempio dati di imaging di risonanza magnetica).

Bibliografia

- [1] Dobson, R., Giovannoni, G. (2018). Multiple sclerosis – a review. In *European Journal of Neurology* (Vol. 26, Issue 1, pp. 27–40). Wiley. <https://doi.org/10.1111/ene.13819>
- [2] Walton, C., King, R., Rechtman, L., Kaye, W., Leray, E., Marrie, R. A., Robertson, N., La Rocca, N., Uitdehaag, B., van der Mei, I., Wallin, M., Helme, A., Angood Napier, C., Rijke, N., Baneke, P. (2020). Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition. In *Multiple Sclerosis Journal* (Vol. 26, Issue 14, pp. 1816–1821). SAGE Publications. <https://doi.org/10.1177/1352458520970841>
- [3] Alastair Compston, Alasdair Coles. *Multiple Sclerosis*. Lancet, 2008.
- [4] Paz-Zulueta, M., Parás-Bravo, P., Cantarero-Prieto, D., Blázquez-Fernández, C., Oterino-Durán, A. (2020). A literature review of cost-of-illness studies on the economic burden of multiple sclerosis. In *Multiple Sclerosis and Related Disorders* (Vol. 43, p. 102162). Elsevier BV. <https://doi.org/10.1016/j.msard.2020.102162>
- [5] Kamm, C. P., Uitdehaag, B. M., Polman, C. H. (2014). Multiple Sclerosis: Current Knowledge and Future Outlook. In *European Neurology* (Vol. 72, Issues 3–4, pp. 132–141). S. Karger AG. <https://doi.org/10.1159/000360528>
- [6] Confavreux, C., Vukusic, S. (2006). Natural history of multiple sclerosis: a unifying concept. In *Brain* (Vol. 129, Issue 3, pp. 606–616). Oxford University Press (OUP). <https://doi.org/10.1093/brain/awl007>
- [7] Elian, M., Nightingale, S., Dean, G. (1990). Multiple sclerosis among United Kingdom-born children of immigrants from the Indian subcontinent, Africa and the West Indies. In *Journal of Neurology, Neurosurgery and Psychiatry* (Vol. 53, Issue 10, pp. 906–911). BMJ. <https://doi.org/10.1136/jnnp.53.10.906>

- [8] Wingerchuk, D. M. (2011). Smoking: effects on multiple sclerosis susceptibility and disease progression. In *Therapeutic Advances in Neurological Disorders* (Vol. 5, Issue 1, pp. 13–22). SAGE Publications. <https://doi.org/10.1177/1756285611425694>
- [9] Hernán, M. A., Jick, S. S., Logroscino, G., Olek, M. J., Ascherio, A., Jick, H. (2005). Cigarette smoking and the progression of multiple sclerosis. In *Brain* (Vol. 128, Issue 6, pp. 1461–1465). Oxford University Press (OUP). <https://doi.org/10.1093/brain/awh471>
- [10] Marrie, R. A., Rudick, R., Horwitz, R., Cutter, G., Tyry, T., Campagnolo, D., Vollmer, T. (2010). Vascular comorbidity is associated with more rapid disability progression in multiple sclerosis. In *Neurology* (Vol. 74, Issue 13, pp. 1041–1047). Ovid Technologies (Wolters Kluwer Health). <https://doi.org/10.1212/wnl.0b013e3181d6b125>
- [11] Stys, P. K., Zamponi, G. W., van Minnen, J., Geurts, J. J. G. (2012). Will the real multiple sclerosis please stand up? In *Nature Reviews Neuroscience* (Vol. 13, Issue 7, pp. 507–514). Springer Science and Business Media LLC. <https://doi.org/10.1038/nrn3275>
- [12] McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H.-P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., Van Den Noort, S., Weinshenker, B. Y., Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. In *Annals of Neurology* (Vol. 50, Issue 1, pp. 121–127). Wiley. <https://doi.org/10.1002/ana.1032>
- [13] Barkhof F, Filippi M, Miller DH, et al. Comparison of MR imaging criteria at first presentation to predict conversion to clinically definite multiple sclerosis. *Brain* 1997;120:2059– 2069.
- [14] Tintore´ M, Rovira A, Martí´nez M, et al. Isolated demyelinating syndromes: comparison of different MR imaging criteria to predict conversion to clinically definite multiple sclerosis. *Am J Neuroradiol* 2000;21:702–706.
- [15] Polman, C. H., Reingold, S. C., Edan, G., Filippi, M., Hartung, H.-P., Kappos, L., Lublin, F. D., Metz, L. M., McFarland, H. F., O’Connor, P. W., Sandberg-Wollheim, M., Thompson, A. J., Weinshenker, B. G., Wolinsky,

- J. S. (2005). Diagnostic criteria for multiple sclerosis: 2005 revisions to the “McDonald Criteria.” In *Annals of Neurology* (Vol. 58, Issue 6, pp. 840–846). Wiley. <https://doi.org/10.1002/ana.20703>
- [16] Polman, C. H., Reingold, S. C., Banwell, B., Clanet, M., Cohen, J. A., Filippi, M., Fujihara, K., Havrdova, E., Hutchinson, M., Kappos, L., Lublin, F. D., Montalban, X., O’Connor, P., Sandberg-Wollheim, M., Thompson, A. J., Wau-bant, E., Weinshenker, B., Wolinsky, J. S. (2011). Diagnostic criteria for multiple sclerosis: 2010 Revisions to the McDonald criteria. In *Annals of Neurology* (Vol. 69, Issue 2, pp. 292–302). Wiley. <https://doi.org/10.1002/ana.22366>
- [17] Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Mar-rie, R. A., Miller, A. E., Miller, D. H., Montalban, X., ... Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. In *The Lancet Neurology* (Vol. 17, Issue 2, pp. 162–173). Elsevier BV. [https://doi.org/10.1016/s1474-4422\(17\)30470-2](https://doi.org/10.1016/s1474-4422(17)30470-2)
- [18] Lublin, F. D., Reingold, S. C., Cohen, J. A., Cutter, G. R., Sorensen, P. S., Thompson, A. J., Wolinsky, J. S., Balcer, L. J., Banwell, B., Barkhof, F., Be-bo, B., Calabresi, P. A., Clanet, M., Comi, G., Fox, R. J., Freedman, M. S., Goodman, A. D., Inglese, M., Kappos, L., ... Polman, C. H. (2014). Defining the clinical course of multiple sclerosis: The 2013 revisions. In *Neurology* (Vol. 83, Issue 3, pp. 278–286). Ovid Technologies (Wolters Kluwer Health). <https://doi.org/10.1212/wnl.0000000000000560>
- [19] Lublin, F. D. (2014). New Multiple Sclerosis Phenotypic Classification. In *European Neurology* (Vol. 72, Issue s1, pp. 1–5). S. Karger AG. <https://doi.org/10.1159/000367614>
- [20] Kantarci, O. H. (2019). Phases and Phenotypes of Multiple Sclerosis. In *CONTINUUM: Lifelong Learning in Neurology* (Vol. 25, Issue 3, pp. 636–654). Ovid Technologies (Wolters Kluwer Health). <https://doi.org/10.1212/con.0000000000000737>
- [21] https://www.aism.it/sclerosi_multipla_forme (ultimo accesso: 12/07/2022)
- [22] Huang, W.-J., Chen, W.-W., Zhang, X. (2017). Multiple sclerosis: Pathology, diagnosis and treatments. In *Experimental and Therapeu-*

- tic Medicine (Vol. 13, Issue 6, pp. 3163–3166). Spandidos Publications. <https://doi.org/10.3892/etm.2017.4410>
- [23] Marivin M. Goldenberg. Multiple Sclerosis Review. Pharmacy and Therapeutics, MediMedia USA, March 2012
- [24] Torkildsen, Ø., Myhr, K. -M., Bø, L. (2015). Disease-modifying treatments for multiple sclerosis – a review of approved medications. In *European Journal of Neurology* (Vol. 23, Issue S1, pp. 18–27). Wiley. <https://doi.org/10.1111/ene.12883>
- [25] Law, M. T., Traboulsee, A. L., Li, D. K., Carruthers, R. L., Freedman, M. S., Kolind, S. H., Tam, R. (2019). Machine learning in secondary progressive multiple sclerosis: an improved predictive model for short-term disability progression. In *Multiple Sclerosis Journal - Experimental, Translational and Clinical* (Vol. 5, Issue 4, p. 205521731988598). SAGE Publications. <https://doi.org/10.1177/2055217319885983>
- [26] Seccia, R., Gammelli, D., Dominici, F., Romano, S., Landi, A. C., Salvetti, M., Tacchella, A., Zaccaria, A., Crisanti, A., Grassi, F., Palagi, L. (2020). Considering patient clinical history impacts performance of machine learning models in predicting course of multiple sclerosis. In F. Lolli (Ed.), *PLOS ONE* (Vol. 15, Issue 3, p. e0230219). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0230219>
- [27] Rocca, M. A., Valsasina, P., Meani, A., Pagani, E., Cordani, C., Cervellin, C., Filippi, M. (2021). Network Damage Predicts Clinical Worsening in Multiple Sclerosis. In *Neurology - Neuroimmunology Neuroinflammation* (Vol. 8, Issue 4, p. e1006). Ovid Technologies (Wolters Kluwer Health). <https://doi.org/10.1212/nxi.0000000000001006>
- [28] Zhao, Y., Wang, T., Bove, R., Cree, B., Henry, R., Lokhande, H., Polgar-Turcsanyi, M., Anderson, M., Bakshi, R., Weiner, H. L., Chitnis, T. (2020). Ensemble learning predicts multiple sclerosis disease course in the SUMMIT study. In *npj Digital Medicine* (Vol. 3, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41746-020-00338-8>
- [29] A. Ibrahim, A. Küderle, H. Gaßner, J. Klucken, B. M. Eskofier, e F. Kluge, «Inertial sensor-based gait parameters reflect patient-reported fatigue in multiple sclerosis.», *J. Neuroengineering Rehabil.*, vol. 17, n. 1, pag. 165, dic. 2020, doi: 10.1186/s12984-020-00798-9.

- [30] R. Bergamaschi et al., «BREMSO: a simple score to predict early the natural course of multiple sclerosis.», *Eur. J. Neurol.*, vol. 22, n. 6, pagg. 981–989, giu. 2015, doi: 10.1111/ene.12696.
- [31] C. C. F. Vasconcelos, J. C. K. Aurenção, R. M. P. Alvarenga, e L. C. S. Thuler, «Long-term MS secondary progression: Derivation and validation of a clinical risk score.», *Clin. Neurol. Neurosurg.*, vol. 194, pag. 105792, lug. 2020, doi: 10.1016/j.clineuro.2020.105792.
- [32] A. Manouchehrinia et al., «Predicting risk of secondary progression in multiple sclerosis: A nomogram.», *Mult. Scler. Houndmills Basingstoke Engl.*, vol. 25, n. 8, pagg. 1102–1112, lug. 2019, doi: 10.1177/1352458518783667.
- [33] A. I. Pisani, A. Scalfari, F. Crescenzo, C. Romualdi, e M. Calabrese, «A novel prognostic score to assess the risk of progression in relapsing-remitting multiple sclerosis patients.», *Eur. J. Neurol.*, vol. 28, n. 8, pagg. 2503–2512, ago. 2021, doi: 10.1111/ene.14859.
- [34] Kurtzke JF, Berlin L. The effects of isoniazid on patients with multiple sclerosis: preliminary report. *AM Rev Tuberc* 1954;70:577-92.
- [35] Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: An expanded disability status scale (EDSS). In *Neurology* (Vol. 33, Issue 11, pp. 1444–1444). Ovid Technologies (Wolters Kluwer Health). <https://doi.org/10.1212/wnl.33.11.1444>
- [36] Zhao, Y., Healy, B. C., Rotstein, D., Guttmann, C. R. G., Bakshi, R., Weiner, H. L., Brodley, C. E., Chitnis, T. (2017). Exploration of machine learning techniques in predicting multiple sclerosis disease course. In S. V. Ramagopalan (Ed.), *PLOS ONE* (Vol. 12, Issue 4, p. e0174866). Public Library of Science (PLoS). <https://doi.org/10.1371/journal.pone.0174866>
- [37] Pinto, M. F., Oliveira, H., Batista, S., Cruz, L., Pinto, M., Correia, I., Martins, P., Teixeira, C. (2020). Prediction of disease progression and outcomes in multiple sclerosis with machine learning. In *Scientific Reports* (Vol. 10, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1038/s41598-020-78212-6>
- [38] Tommasin, S., Coccozza, S., Taloni, A., Gianni, C., Petsas, N., Pontillo, G., Petracca, M., Ruggieri, S., De Giglio, L., Pozzilli, C., Brunetti, A., Pantano, P. (2021). Machine learning classifier to identify clinical and radiological features

- relevant to disability progression in multiple sclerosis. In *Journal of Neurology* (Vol. 268, Issue 12, pp. 4834–4845). Springer Science and Business Media LLC. <https://doi.org/10.1007/s00415-021-10605-7>
- [39] Yperman, J., Becker, T., Valkenburg, D., Popescu, V., Hellings, N., Wijmeersch, B. V., Peeters, L. M. (2020). Machine learning analysis of motor evoked potential time series to predict disability progression in multiple sclerosis. In *BMC Neurology* (Vol. 20, Issue 1). Springer Science and Business Media LLC. <https://doi.org/10.1186/s12883-020-01672-w>
- [40] LaRocca, N. G., Hudson, L. D., Rudick, R., Amtmann, D., Balcer, L., Benedict, R., Bermel, R., Chang, I., Chiaravalloti, N. D., Chin, P., Cohen, J. A., Cutter, G. R., Davis, M. D., DeLuca, J., Feys, P., Francis, G., Goldman, M. D., Hartley, E., . . . Kapoor, R. (2017). The MSOAC approach to developing performance outcomes to measure and monitor multiple sclerosis disability. In *Multiple Sclerosis Journal* (Vol. 24, Issue 11, pp. 1469–1484). SAGE Publications. <https://doi.org/10.1177/1352458517723718>
- [41] Rudick, R. A., LaRocca, N., Hudson, L. D. (2013). Multiple Sclerosis Outcome Assessments Consortium: Genesis and initial project plan. In *Multiple Sclerosis Journal* (Vol. 20, Issue 1, pp. 12–17). SAGE Publications. <https://doi.org/10.1177/1352458513503392>
- [42] Goldman, M. D., LaRocca, N. G., Rudick, R. A., Hudson, L. D., Chin, P. S., Francis, G. S., Jacobs, A., Kapoor, R., Matthews, P. M., Mowry, E. M., Balcer, L. J., Panzara, M., Phillips, G., Uitdehaag, B. M. J., Cohen, J. A. (2019). Evaluation of multiple sclerosis disability outcome measures using pooled clinical trial data. In *Neurology* (Vol. 93, Issue 21, pp. e1921–e1931). Ovid Technologies (Wolters Kluwer Health). <https://doi.org/10.1212/wnl.00000000000008519>
- [43] Hernandez, L., O'Donnell, M., Postma, M. (2020). Predictors of Health Utility in Relapsing–Remitting and Secondary-Progressive Multiple Sclerosis: Implications for Future Economic Models of Disease-Modifying Therapies. In *Pharmacoeconomics* (Vol. 39, Issue 2, pp. 243–256). Springer Science and Business Media LLC. <https://doi.org/10.1007/s40273-020-00964-w>
- [44] Cohen, J. A., Reingold, S. C., Polman, C. H., Wolinsky, J. S. (2012). Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. In *The Lancet Neurology* (Vol. 11, Issue 5, pp. 467–476). Elsevier BV. [https://doi.org/10.1016/s1474-4422\(12\)70059-5](https://doi.org/10.1016/s1474-4422(12)70059-5)

- [45] Motl, R. W., Cohen, J. A., Benedict, R., Phillips, G., LaRocca, N., Hudson, L. D., Rudick, R. (2017). Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. In *Multiple Sclerosis Journal* (Vol. 23, Issue 5, pp. 704–710). SAGE Publications. <https://doi.org/10.1177/1352458517690823>
- [46] Benedict, R. H., DeLuca, J., Phillips, G., LaRocca, N., Hudson, L. D., Rudick, R. (2017). Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. In *Multiple Sclerosis Journal* (Vol. 23, Issue 5, pp. 721–733). SAGE Publications. <https://doi.org/10.1177/1352458517690821>
- [47] [https://www.nationalmssociety.org/For-Professionals/Researchers/Resources-for-MS-Researchers/Research-Tools/Clinical-Study-Measures/Paced-Auditory-Serial-Addition-Test-\(PASAT\)](https://www.nationalmssociety.org/For-Professionals/Researchers/Resources-for-MS-Researchers/Research-Tools/Clinical-Study-Measures/Paced-Auditory-Serial-Addition-Test-(PASAT)) (ultimo accesso: 12/07/2022)
- [48] TOMBAUGH, T. (2006). A comprehensive review of the Paced Auditory Serial Addition Test (PASAT). In *Archives of Clinical Neuropsychology* (Vol. 21, Issue 1, pp. 53–76). Oxford University Press (OUP). <https://doi.org/10.1016/j.acn.2005.07.006>
- [49] Lamers, I., Cattaneo, D., Chen, C. C., Bertoni, R., Van Wijmeersch, B., Feys, P. (2015). Associations of Upper Limb Disability Measures on Different Levels of the International Classification of Functioning, Disability and Health in People With Multiple Sclerosis. In *Physical Therapy* (Vol. 95, Issue 1, pp. 65–75). Oxford University Press (OUP). <https://doi.org/10.2522/ptj.20130588>
- [50] Feys, P., Lamers, I., Francis, G., Benedict, R., Phillips, G., LaRocca, N., Hudson, L. D., Rudick, R. (2017). The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. In *Multiple Sclerosis Journal* (Vol. 23, Issue 5, pp. 711–720). SAGE Publications. <https://doi.org/10.1177/1352458517690824>
- [51] Shahid, A., Wilkinson, K., Marcu, S., Shapiro, C. M. (2011). Beck Depression Inventory. In *STOP, THAT and One Hundred Other Sleep Scales* (pp. 63–64). Springer New York. https://doi.org/10.1007/978-1-4419-9893-4_8
- [52] E Rouillet, M Coste, V Basque, D Moccatti, C Beneton, M F Millet, R Marteau. [The Minimal Record of Disability and multiple sclerosis. Results of the evaluation of 200 patients]. 1988;144(12):805-16. PMID: 3070693

- [53] Hays, R. D., Morales, L. S. (2001). The RAND-36 measure of health-related quality of life. In *Annals of Medicine* (Vol. 33, Issue 5, pp. 350–357). Informa UK Limited. <https://doi.org/10.3109/07853890109002089>
- [54] Hays, R. D., Sherbourne, C. D., Mazel, R. M. (1993). The rand 36-item health survey 1.0. In *Health Economics* (Vol. 2, Issue 3, pp. 217–227). Wiley. <https://doi.org/10.1002/hec.4730020305>
- [55] Busija, L., Pausenberger, E., Haines, T. P., Haymes, S., Buchbinder, R., Osborne, R. H. (2011). Adult measures of general health and health-related quality of life: Medical Outcomes Study Short Form 36-Item (SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Sh. In *Arthritis Care amp; Research* (Vol. 63, Issue S11, pp. S383–S412). Wiley. <https://doi.org/10.1002/acr.20541>
- [56] Balcer, L. J., Raynowska, J., Nolan, R., Galetta, S. L., Kapoor, R., Benedict, R., Phillips, G., LaRocca, N., Hudson, L., Rudick, R. (2017). Validity of low-contrast letter acuity as a visual performance outcome measure for multiple sclerosis. In *Multiple Sclerosis Journal* (Vol. 23, Issue 5, pp. 734–747). SAGE Publications. <https://doi.org/10.1177/1352458517690822>
- [57] Azur, M. J., Stuart, E. A., Frangakis, C., Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? In *International Journal of Methods in Psychiatric Research* (Vol. 20, Issue 1, pp. 40–49). Wiley. <https://doi.org/10.1002/mpr.329>
- [58] Longadge, R., Dongre, S. (2013). Class Imbalance Problem in Data Mining Review (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.1305.1707>
- [59] Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., Napolitano, A. (2010). RU-SBoost: A Hybrid Approach to Alleviating Class Imbalance. In *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* (Vol. 40, Issue 1, pp. 185–197). Institute of Electrical and Electronics Engineers (IEEE). <https://doi.org/10.1109/tsmca.2009.2029559>