

Il Paradigma di Gemello Digitale - *Digital Twin*

Sergio Cerutti DEIB - Politecnico di Milano, Presidente del Comitato Etico dell'Istituto Europeo di Oncologia - IEO e Centro Cardiologico "Monzino", Milano

In questo lavoro si delinea il concetto di *Digital Twin* e verranno messi in evidenza gli aspetti più innovativi dell'elaborazione dei dati e dei segnali biomedici che permettono sin da ora di ricavare informazioni, impiegando diversi canali di informazione e diverse modalità di rappresentazione

Non vi è dubbio che la Medicina stia diventando sempre più "quantitativa": ciò è dovuto anche all'enorme progresso delle Tecnologie Biomediche che hanno permesso negli ultimi decenni di poter ottenere parametri sempre più precisi e "personalizzati" per le varie decisioni cliniche (diagnosi, terapia e riabilitazione). Oggi certamente possiamo realizzare dei file personali dei pazienti in cui memorizzare tutte le informazioni anagrafiche, anamnestiche e di decorso, potendo così avere un facile accesso ad una mole enorme di dati che è certamente di estrema importanza per valutare la storia clinica del soggetto e quindi per intervenire in caso di decisioni diagnostiche e/o terapeutiche. Attualmente, molte Regioni nel nostro Paese hanno sviluppato un Fascicolo Sanitario Elettronico (FSE), dove far confluire tutte le informazioni che possano essere di interesse dal punto di vista sanitario e clinico. Dato l'enorme sviluppo delle tecnologie informatiche anche in questo contesto applicativo, si sta delineando il concetto di *Digital Twin* [1-2] o Gemello Digitale, inteso come quel modello "antro-

pomorfo" che si potrebbe implementare con tutte le informazioni sopra menzionate, compresi tutti gli esami di laboratorio, le registrazioni di segnali e immagini che potrebbero in qualche modo "aggiornare" l'informazione già presente. L'idea sarebbe quella di creare un "alter-ego" su calcolatore, un modello del singolo soggetto, con le informazioni che di volta in volta, nel corso della sua vita, verranno disponibili. Quindi questo "Gemello Digitale" dovrebbe invecchiare anche lui, per poter essere sempre aggiornato ed essere in grado di descrivere il più possibile fedelmente e in tempo reale la situazione clinica del soggetto. È evidente che siamo ancora lontani dal poter implementare un modello di questo genere, ma è certamente importante avere in mente tale obiettivo da realizzare magari nei prossimi decenni.

In questo contesto, gioca un ruolo fondamentale il tipo di informazione e le modalità con cui questa viene fornita al calcolatore. In questo lavoro verranno messi in evidenza gli aspetti più innovativi dell'elaborazione dei dati e dei segnali biomedici che permettono sin da ora di ricavare informazioni, impiegando diversi canali di informazione e diverse modalità di rappresentazione. Viene ribadito il concetto di "Paradigma 4-M" (Multivariato, Multiorgano, Multimodale e Multiscala) che rappresenta un passo in avanti verso quella integrazione dei dati, oggi sempre più richiesta sia negli studi fisiologici che nella pratica clinica.

Verso il Concetto di *Digital Twin*

Mediante le sempre più diffuse tecnologie avanzate nel settore biomedico, si è attualmente in grado di poter realizzare, almeno concettualmente, un *alter-ego* del singolo paziente e di memorizzare le relative informazioni in un apposito dispositivo informatico "in silico", vale a dire su supporto collegato ad uno o più calcolatori elet-


tronici. Su tale supporto si possono inserire molteplici e diverse informazioni relative al paziente, in modo da realizzare una “simulazione” il più possibile realistica del paziente: una specie di “gemello digitale” (o *Digital Twin*) [1-4], con il vantaggio di poter effettuare studi integrati, coinvolgendo i vari organi del corpo umano coinvolti, e operare interventi “virtuali” sul modello del paziente, valutando così l’effetto di una terapia rispetto ad un’altra. Ovviamente il concetto di *Digital Twin* costituirebbe un punto di arrivo (certamente ancora non completamente realizzato) perché ci sarà sempre una differenza, anche molto significativa, tra paziente effettivo e il suo modello digitale. Negli ultimi anni, comunque, si sono fatti notevoli progressi nel settore dell’integrazione di da-

ti, segnali e immagini, per cui un grande numero di informazioni utili per la fase diagnostica e terapeutica possono essere intelligentemente fornite al calcolatore e da qui ricavare, mediante algoritmi opportuni di classificazione e di decisione, suggerimenti e indicazioni cliniche preziose e difficilmente ottenibili con approcci diversi. La figura 1 rappresenta, in maniera molto olistica, come si possa riversare informazione utile dal paziente, nella sua unicità, su un modello e quindi su un supporto informatico di esso. Mentre la figura 2 mostra le enormi potenzialità di integrazione tra dati, segnali, immagini e altra informazione per ottenere un archivio (*repository di dati*) da cui estrarre informazione utile da un punto di vista medico. Non c’è dubbio che le molteplicità di modi



Figura 1

La prospettiva di un Gemello Digitale, ottenuto “in silico”, sulla base di una opportuna integrazione di dati, segnali e immagini ottenuti dal singolo paziente

Figura 2  *La memorizzazione di un archivio il più possibile completo e personalizzato di dati relativi a un singolo paziente per poter ottenere informazioni integrate e “arricchite” da algoritmi di Intelligenza Artificiale*



con cui viene documentata questa informazione può trovare giovamento dall'applicazione di tecniche avanzate di Intelligenza Artificiale [5] che permettono di ottenere inferenze statistiche significative da dati di origine diversa (anamnestici, di esame obiettivo, di espressione di segni e sintomi, di risultati ottenuti con terapie condotte in precedenza). Le notevoli potenzialità espresse da strumenti di calcolo e di elaborazione sempre più complessi e sofisticati, uniti ad algoritmi tipici dell'Intelligenza Artificiale (IA), permettono di rilevare correlazioni e inferenze statistiche su una mole notevole di dati, peraltro anche assai disomogenei [6]. Particolare attenzione deve essere però dedicata per l'elaborazione della struttura di dati ottimale che possa determinare delle relazioni causali stabili e davvero significative, per non incorrere nel tipico errore del *garbage in, garbage out*: vale a dire, dati non ben articolati in ingresso non possono che determinare dati con incredibili errori in uscita, molte volte neanche misurabili correttamente. Insomma, bisogna avere le competenze appropriate per apprezzare le giuste potenzialità degli algoritmi di IA, senza credere quasi in maniera "mistica", ma erronea, che tali algoritmi siano sempre in grado di ottenere risultati statisticamente significativi. Il concetto di "diagnosi clinica" è certamente molto più complesso di una struttura ramificata ad albero in cui si deve individuare il percorso più probabile. Molte volte la patologia deriva proprio da un errore statistico, in cui può accadere che sia proprio la struttura meno probabile quella che si manifesta in realtà.

Paradigma 4-M (Multivariato, Multimodale, Multiorgano e Multiscala) nell'elaborazione avanzata di segnali biomedici

Recentemente [7-9] è stato introdotto il concetto di "Paradigma 4-M": Multivariato, Multimodale, Multiorgano e Multiscala, in una visione più moderna e avanzata delle tecniche di analisi ed elaborazione di segnali e immagini biomediche. Il prefisso "multi" attira la nostra attenzione sul fatto che l'informazione viaggia su vari segnali (con più derivazioni) e non solo: segnali diversi presentano delle caratteristiche diverse che possono essere messe insieme (integrate) in diverse rappresentazioni. Nel seguito verranno esaminate varie implementazioni di questo paradigma, con semplici applicazioni ai segnali e alle immagini biomediche.

Approccio Multivariato

Tale approccio consiste nell'utilizzazione di vari canali rilevati dalla stessa sorgente di dati. Come esempio, si può considerare la registrazione dei

segnali EEG (elettroencefalografici) sulla teca cranica di un soggetto. Il tradizionale *montage* impiegato per usi clinici, utilizza 21 elettrodi distribuiti sulla teca cranica e sul padiglione auricolare/mastoide (sistema internazionale 10-20). Ovvero con un numero ancora più elevato di elettrodi (128 o 256) per un'analisi detta "*in alta risoluzione spaziale*". Dati k canali EEG, si può ottenere un modello autoregressivo (AR) multivariato dell'intero fenomeno elettrico (Figura 3):

$$X(t) = (X_1(t), X_2(t), \dots, X_k(t))$$

$$X(t) = \sum_{j=1}^p A(j)X(t-j) + E(t)$$

Dove $A(0) = (-1)$: matrice di identità e

$$E(t) = \sum_{j=0}^p A(j)X(t-j).$$

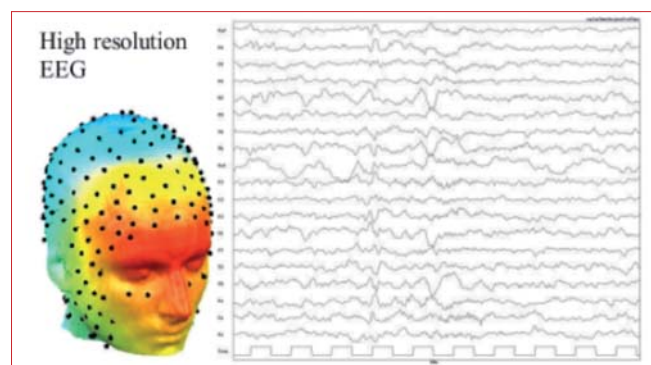
Passando dalla trasformata-z, si ricava:

$$X(f) = A^{-1}(f)E(f) = H(f)E(f).$$

In $H(f)$ (matrice di trasferimento), viene contenuta tutta l'informazione delle auto e cross-correlazioni tra i canali che possono essere così opportunamente calcolate.

Approccio Multimodale

Con questo approccio, si vuole ottenere una integrazione dell'informazione sullo stesso sistema fisiologico, ma impiegando diverse "modalità". Ad esempio, la figura 4a rappresenta una immagine tridimensionale di Risonanza Magnetica con la visualizzazione della posizione di elettrodi realizzati per essere visibili in Risonanza, per una registrazione simultanea di attività elettrica cerebrale (EEG). Vengono anche visualizzate le linee di forza del campo elettrico generato, ottenute collegando tra loro punti con uguale valore di potenziale elettrico. La figura 4b mostra una possibile risoluzione inversa del campo elettrico, con indi-



▲ Figura 3

Registrazione multicanale e in alta risoluzione spaziale del segnale EEG, come esempio di analisi Multivariata. A titolo di esempio si mostrano i canali del tradizionale montage 10-20

cata in rosso la localizzazione della sorgente. Si possono inoltre compiere operazioni più sofisticate, come ad esempio il poter passare da una visualizzazione “tradizionale” del segnale di Risonanza e dell’EEG sulla teca cranica, ad una visualizzazione che permetta l’“eliminazione” dell’osso cranico e dei foglietti superficiali della massa cerebrale, mettendo in evidenza i primi strati della corteccia cerebrale sia in Risonanza che nel calcolo della distribuzione del potenziale (Figura 4c)

Approccio Multiorgano

Molteplici sono le situazioni fisiopatologiche in cui si osserva chiaramente una interazione tra organi diversi (Sistema Nervoso Centrale, Sistema Autonomo, Sistema Cardiovascolare, Sistema Neuromuscolare, Sistema Endocrino-Metabolico, ecc.). Vi sono alcune patologie, in Clinica Medica, che si possono considerare “Patologie Multiorgano”. Ad esempio, l’alterazione delle varie fasi del sonno in caso di patologie centrali, lo svilupparsi di apnee notturne, varie disfunzioni autonome e molte altre, sono caratterizzate da compromissioni che vedono coinvolti più sistemi fisiologici contemporaneamente e dove questa interazione risulta di fondamentale importanza per la comprensione corretta delle cause che le hanno provocate.

Come “studio di caso” di questo approccio “Multiorgano” è stata scelta la Patologia così detta RLLS (*Rest-Less-Leg-Syndrome*) - *Sindrome delle Gambe Senza Riposo*. La figura 5, solo a titolo di esempio, riporta il diretto coinvolgimento di questa sindrome verso 4 compartimenti fisiologici: il Sistema Nervoso Centrale, il Sistema Nervoso Autonomo, il Sistema Respiratorio e il Sistema Neuromuscolare. Questi 4 Sistemi presentano comportamenti significativi (e diversi dalla situazione di normalità) durante episodi di RLLS. Tale Sindrome, significativamente complessa, è detta anche di Willis-Ekbom ed è un disturbo neurologico che determina movimenti parossistici e senza controllo (generalmente degli arti inferiori) che si manifesta prevalentemente a riposo in posizione supina o anche in determinate fasi del sonno [10].

Nella figura 5 vengono riportati 4 segnali, relativi ai Sistemi sopra indicati, che risultano i seguenti: segnale EEG, segnale di Variabilità Cardiaca (HRV), Respiro e segnale EMG (dal muscolo peroneale della gamba). L’epifenomeno principale che si manifesta con tale patologia è la contrazione più o meno regolare e sincronizzata del muscolo della gamba che è chiaramente correlata con *burst* di attività centrale visibile sul segnale EEG.

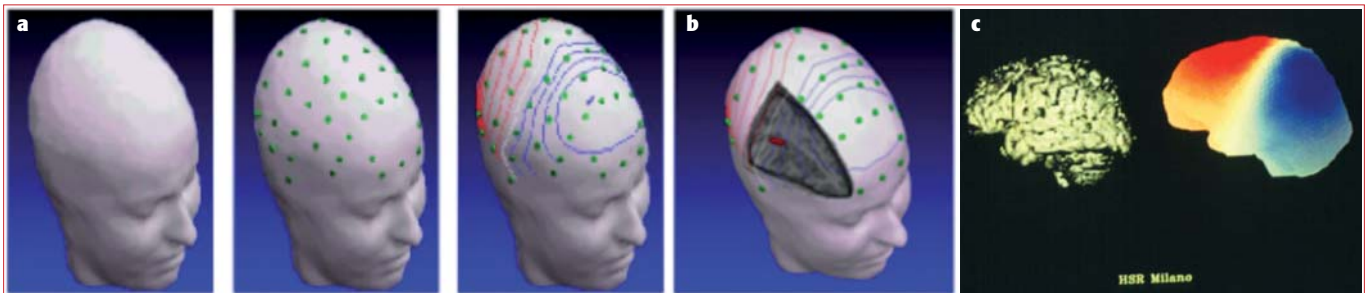


Figura 4
a Esempio di un approccio Multimodale: fusione di Risonanza Magnetica con elettrodi EEG (montage 10-20); **b** Localizzazione delle sorgenti, tramite risoluzione del problema inverso; **c** “rimozione” in silico della teca cranica e dei foglietti superficiali esterni per ottenere la distribuzione del potenziale a livello della corteccia cerebrale

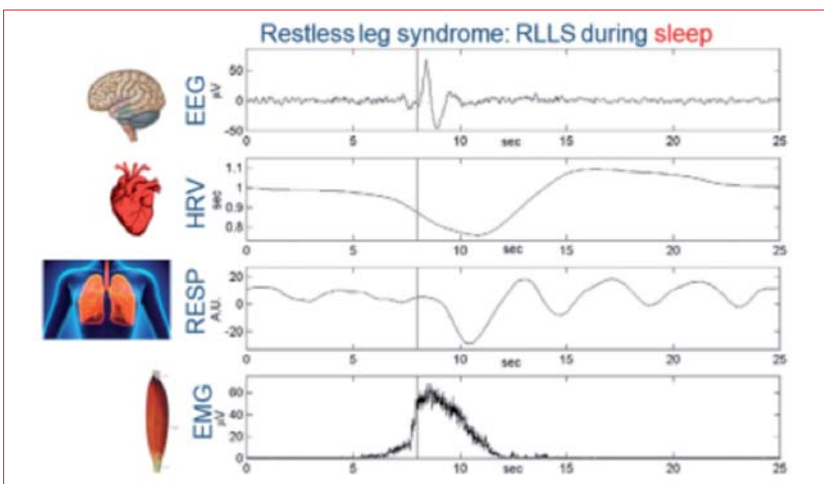


Figura 5
 Approccio Multiorgano. Nel caso della *Sindrome delle Gambe senza Riposo (RLLS)* si nota una sincronizzazione di 4 segnali biologici: EEG, HRV, respiro e segnale EMG. Calcolando la media sincrona dei 4 segnali avendo come trigger lo spike di origine centrale sull’EEG, si ottengono delle risposte coerenti degli altri segnali a cui si possono attribuire delle valenze di tipo clinico

Calcolando la media sincronizzata dei vari segnali sull'istante di occorrenza di tali *burst*, si può verificare che esiste una chiara correlazione, almeno temporale. Infatti, in corrispondenza della contrazione dell'arto abbiamo una tachicardia (diminuzione dell'HRV), un respiro che diventa più regolare - forse dovuto al fatto che il soggetto "sente" le contrazioni involontarie -, e infine un'attività elettromiografica in concomitanza con il *burst* EEG. Anche altri parametri risultano "sincronizzati", come ad esempio parametri metabolici che qui non vengono indicati. La Sindrome RLLS risulta interessante perché con chiari e prevedibili comportamenti patologici di alcuni sistemi sottintende un più ampio coinvolgimento di una molteplicità di sistemi: insomma, una "tempesta perfetta" che coinvolge parecchi sistemi e funzionalità biologiche, la cui patogenesi non è ancora completamente nota.

Approccio Multiscala

L'approccio "Multiscala" ha lo scopo di integrare l'informazione, collegando scale diverse di osservazione del fenomeno biologico in studio. Come si può valutare nella figura 6, con le attuali conoscenze sviluppate per le tecniche di studio del genoma e del proteoma [11-12], è possibile individuare geni specifici che sovrintendono alla costruzione di proteine. Strutture

proteiche più o meno complesse, danno luogo a cellule che sono specializzate a comportarsi in un certo modo. L'insieme di cellule costituisce il tessuto biologico e da qui si passa all'organo nella sua interezza. È facile a questo punto comprendere che qualche "malfunzionamento" o "errore di trascrizione" (o altro) nella catena di sviluppo biologico, passando dal DNA della cellula, fino ad arrivare all'organo in esame, possa causare qualche "anormalità" nel consueto e normale sviluppo.

Tali anomalie possono generare, a seconda dei contesti, situazioni patologiche gravi, meno gravi, del tutto trascurabili o addirittura neanche individuabili con tecniche tradizionali nella clinica attuale.

In questo scenario, appare senz'altro interessante e molto promettente per il futuro, il poter seguire efficacemente questo per-

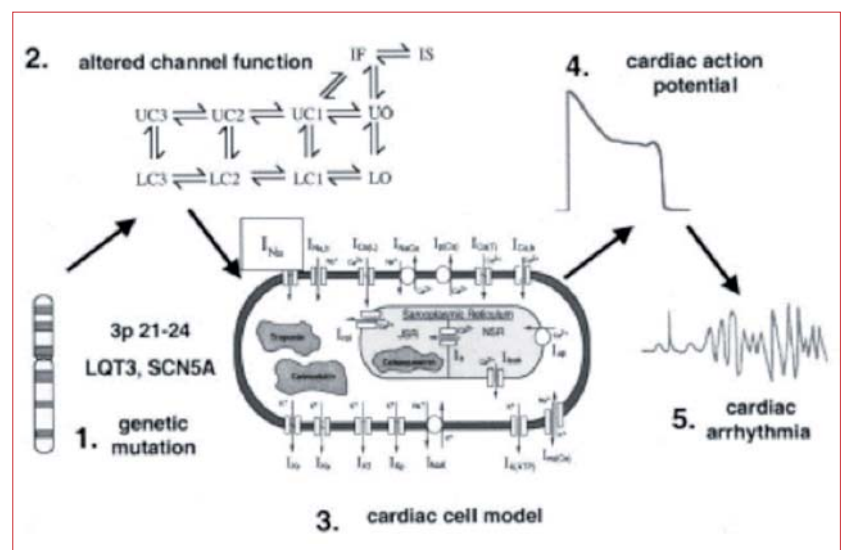


Figura 7

Come una mutazione cromosomica (1) (sul locus 3p 21-24) può provocare un allungamento del tratto QT sull'ECG (sindrome del QT lungo) (4). La mutazione (1) causa una alterazione delle correnti di canale (principalmente Na⁺, K⁺ e Cl⁻) (2), sulla base di un modello (3) che determina l'allungamento del tratto QT (4) che è correlato con lo svilupparsi di gravi aritmie (5). Da [13]

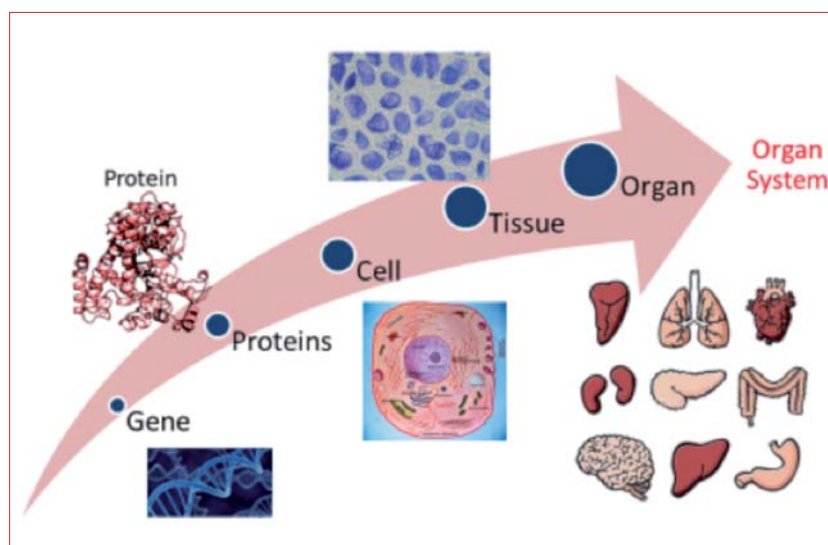


Figura 6

Passaggio Multiscala: dalle sequenze geniche, alla formazione di proteine, di cellule e di tessuti per arrivare sino alla scala dell'organo completo

corso “multiscala” (dalla sequenza genica, allo sviluppo dell’intero organo ovvero della sua corretta funzionalità) e valutare se ci sono stati alcuni “intoppi” o “non conformità” rispetto ad una situazione normale.

Un esempio maggiormente chiarificatore di questa complessa relazione geni/organo/sua funzionalità viene illustrato nella figura 7 da Y. Rudy e riportata da [13]. Rudy ha sviluppato un modello dove ha potuto dimostrare che la “sindrome del QT-lungo” non sia altro che un epifenomeno di una cascata più complessa di eventi che partono dalla sequenza genica, fino alla generazione del potenziale di azione a livello ventricolare (depolarizzazione ventricolare). Partendo quindi da una sequenza di DNA mutata, si passa a considerare la variazione indotta da tale mutazione sull’equilibrio degli ioni che determinano le correnti che passano attraverso la membrana cellulare. Il modello di Rudy tiene conto infatti delle correnti Na, K e Cl. Dall’andamento di tali correnti viene generata la differenza di potenziale all’interno delle cellule ventricolari, rappresentata dal potenziale di azione. Supponiamo che nel corredo genetico di un individuo vi sia una mutazione, individuata come 3p 21-24. Tale mutazione dà origine ad una forma patologica chiamata QT-lungo che si evidenzia facilmente sul segnale elettrocardiografico (ECG) come allungamento del tratto QT, rispetto ad una situazione normale di riferimento. Pazienti con QT-lungo presentano maggiori probabilità di sviluppare aritmie ed altri seri disturbi elettrocardiografici e non solo [14].

Diverse sono oggi le conoscenze che possono indicare una origine genetica a varie patologie: alcune forme di ipertensione, di epilessia, la fibrosi cistica, ecc. Risulta quindi di fondamentale importanza poter collegare tra loro le diverse scale di osservazione e valutare le origini di sintomi e segni caratterizzanti.

Conclusioni

È stato compiuto un viaggio, mi auguro interessante e stimolante, all’interno delle applicazioni di tecniche avanzate di elaborazione dei dati e dei segnali biomedici. Il paradigma 4M – Multivariato, Multiorgano, Multimodale e Multiscala fornisce un riferimento delle potenzialità che l’integrazione opportuna di queste informazioni può fornire in chiave diagnostica al medico per operare le sue decisioni. È certamente suggestivo poter immaginare un “Gemello Digitale”, realizzato *in silico*, sulla base di tutti questi dati che possa descrivere il comportamento del singolo paziente, la risposta a possibili terapie e simulare quindi un atto medico o chirurgico, integrando in

modo intelligente tutti i dati ricavabili. Alcuni strumenti di calcolo e d’interpretazione sono già attualmente disponibili, ma certamente la strada è ancora molto lunga per implementare, anche in minima parte, questa visione così ambiziosa ma allo stesso tempo affascinante.

BIBLIOGRAFIA

- [1] K. Bruynseels, F. Santoni de Sio, J. van den Hoven: Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm, *Front Genet.*, 13 febbraio 2018.
- [2] A. Croatti, M. Gabellini, S. Montagna, A. Ricci: On the Integration of Agents and Digital Twins, *Healthcare, J Med Syst.*, 4 agosto 2020.
- [3] C. Angulo, L. Gonzalez-Abril, C. Raya, J.A. Ortega: A Proposal to Evolving Towards Digital Twins in Healthcare, *Lecture Notes in Computer Science*, 2020, pp. 418-426.
- [4] C.G. Leo, M.R. Tumolo, S. Sabina, R. Colella, V. Recchia, G. Ponzini, D.I. Fotiadis, A. Bodini, P. Mincaroni: Health Technology Assessment for In Silico Medicine: Social, Ethical and Legal Aspects, *International Journal of Environmental Research and Public Health*, 19(3), 2022.
- [5] P. Shah, F. Kendall, S. Khozin et al.: Artificial intelligence and machine learning in clinical development: a translational perspective, *npj Digit. Med.*, 69(2), 2019.
- [6] R. Dias, A. Torkamani: Artificial intelligence in clinical and genomic diagnostics, *Genome Med.*, 11(1):70, 19 novembre 2019.
- [7] S. Cerutti: Multivariate, Multiorgan and Multiscale Integration of Information in Biomedical Signal Processing, *IET 3rd International Conference on Advances in Medical Signals and Information Processing-MEDSIP*, Glasgow, UK, 2006.
- [8] S. Cerutti, D. Hoyer, A. Voss: Multiscale, Multiorgan and Multivariate complexity analyses of cardiovascular Regulation, *Phil. Trans. R. Soc. A.*, 367, 2009, pp. 1337-1358.
- [9] S. Cerutti: Multivariate and multiscale analysis of biomedical signals: Towards a comprehensive approach to medical diagnosis, *25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, 2012, pp. 1-5.
- [10] D. Garcia-Borreguero, R. Kohnen, M.H. Silber, J.W. Winkelman, C.J. Earley, B. Högl, M. Manconi, J. Montplaisir, Y. Inoue, R.P. Allen: The long-term treatment of restless legs syndrome/Willis-Ekbom disease: evidence-based guidelines and clinical consensus best practice guidance: a report from the International Restless Legs Syndrome Study Group, *Sleep Med.*, 14(7), 14 luglio 2013, pp. 675-684.
- [11] C.H. Luo, Y. Rudy: A dynamic model of the cardiac ventricular action potential. I. Simulations of ionic currents and concentration changes, *Circulation Research*, 74, 1994, pp. 1071-1097.
- [12] D. Noble: Modeling the Heart: from Genes to Cells to the Whole Organ, *Science*, vol. 295, 1 marzo 2002, pp. 1668-1682, 2002.
- [13] Y. Rudy: From Genome to Physiome: Integrative Models of Cardiac Excitation, *Annals of Biomedical Engineering*, vol. 28, 2000, pp. 945-950.
- [14] S.R. Shah, K. Park, R. Alweis: Long QT Syndrome: A Comprehensive Review of the Literature and Current Evidence, *Curr Probl Cardiol.*, 44(3), marzo 2019, pp. 92-106.