

Medicina e intelligenza artificiale: come le macchine possono migliorare la nostra salute

Alessandro Rizzato, Data Scientist at Deepclever by Polaris Engineering S.r.l.

TOPIC

Col passare degli anni, sono sempre più le discipline che stanno mostrando un grande interesse nei confronti dell'intelligenza artificiale: una di queste è il settore medico, con la ricerca proiettata nell'utilizzo di tecniche di Machine Learning in grado di migliorare le diagnosi, individuare le schematicità di una ben determinata patologia e, più in generale, di fornire strumenti migliori a disposizione dei dottori, in modo da poter intervenire sui singoli pazienti tramite contromisure adattate ad esigenze specifiche. Non solo: le recenti tecniche di deep learning potrebbero consentire in un futuro prossimo di avere fasi delle terapie totalmente automatizzate, permettendo a medici e operatori del settore di dedicarsi maggiormente allo studio di nuove soluzioni.

Sommario

Motivazioni nell'applicare l'AI in campo medico	2
Prognosi	2
Diagnosi.....	2
Trattamento	2
Flussi di lavoro per i medici.....	2
Coinvolgimento di un maggior numero di esperti	3
Tecniche di Machine Learning comunemente usate in campo medico	3
Support Vector Machine	3
Convolutional Neural Network	4
Random Forest.....	6
Limiti dell'AI in campo medico.....	6
Disponibilità di dati di qualità	6
Privacy	6
Apprendimento da pratiche passate sbagliate	6
Esperienza nella valutazione finale.....	6
Cooperazione interdisciplinare	7
Tecniche di AI applicate al COVID-19.....	7
Conclusioni	8
References.....	8

AI in campo medico

Cosa succederebbe se un gruppo di esperti potesse essere sempre ai servizi di un singolo paziente? Come fare per ridurre gli errori che possono essere compiuti nella formulazione di una diagnosi? In un mondo ideale, ogni terapia verrebbe formulata andando incontro con estrema precisione alle esigenze del singolo, prendendo spunto sia dalla sua storia clinica che dalla conoscenza collettiva data da milioni di altri pazienti: grazie al Machine Learning, i medici possono rompere i limiti imposti da una conoscenza circostanziale, derivante dall'applicazione di terapie su un numero ristretto di pazienti, attraverso l'uso di una “esperienza condivisa”, grazie alla quale formulare terapie ottimizzate per una persona specifica [1]. Non solo: l'adozione di tecniche di intelligenza artificiale può permettere la totale automazione di alcune fasi cliniche, agevolando il lavoro degli specialisti.

Prognosi

Una prognosi è un procedimento grazie al quale si può predire lo sviluppo di una certa malattia. Un modello di machine learning è in grado di permettere ai medici di prevedere eventi futuri: tra quanto verosimilmente un paziente potrà tornare alla vita di tutti i giorni dopo essersi sottoposto a delle cure? Quanto velocemente progredirà una malattia? Un modello algoritmico richiede dei dati in grado di fornire un quadro completo, inclusi i risultati delle terapie alle quali il paziente si è sottoposto in passato [1]. A livello clinico, andranno collezionati diversi tipi di dati, quali: risultati di test fenotipici, genomici, proteomici e patologici, congiuntamente a immagini mediche [2].

Diagnosi

I migliori dottori sono in grado di comprendere quando un particolare avvenimento clinico sia effettivamente nella norma oppure se rappresenti un rischio per la salute del paziente.

L'americana Institute of Medicine ha evidenziato che tutte le persone nel corso della loro vita incorrono almeno una volta in una diagnosi errata [3]: ridurre qualsiasi tipo di errore può essere cruciale nel caso di patologie non comuni, senza considerare il fatto che ciò possa avere un effetto benefico perfino nel caso di malattie a noi più familiari. Basti pensare che complicazioni ritenute debellate quali tubercolosi e/o dissenteria hanno almeno una possibilità di non essere individuate, sebbene nei paesi sviluppati vi sia un accesso adeguato a terapie in grado di far fronte a queste disfunzioni [4]. Attraverso i dati collezionati durante terapie di tutti i giorni, le tecniche di AI possono individuare le diagnosi più probabili durante una visita clinica e prospettare quali condizioni si manifesteranno in futuro nel paziente [1].

Trattamento

In un sistema sanitario su scala nazionale, con migliaia di medici impegnati con altrettanti pazienti, potrebbero presentarsi delle variazioni circa le modalità di trattamento di certi sintomi. Un algoritmo di ML può individuare queste naturali variazioni in modo da aiutare i dottori ad identificare un trattamento da preferire rispetto ad un altro [1]. Un'applicazione potrebbe essere quella di confrontare ciò che il dottore prescriverebbe a un paziente con un trattamento suggerito da un modello algoritmico [1].

Flussi di lavoro per i medici

L'introduzione delle **cartelle elettroniche (EHR, Electornic Health Record)** ha agevolato l'accesso ai dati, ma allo stesso tempo ha fatto emergere dei “colli di bottiglia” derivanti da passaggi burocratici e amministrativi, creando ulteriori complicazioni per i medici. Le tecniche di ML possono permettere lo snellimento di passaggi inefficienti e macchinosi all'interno del flusso di lavoro clinico [1]. Le stesse tecnologie che vengono utilizzate nei motori di ricerca possono porre

in evidenza informazioni rilevanti nella cartella clinica di un paziente, agevolando il lavoro degli specialisti. Ciò consente inoltre ulteriori facilitazioni nell'inserimento di nuovi dati tenendo conto dello storico clinico di un soggetto [1].

Coinvolgimento di un maggior numero di esperti

L'adozione dell'intelligenza artificiale può dare la possibilità di raggiungere un maggior numero di specialisti in grado di fornire una valutazione medica senza che vi sia un loro coinvolgimento diretto [1]. Ad esempio: un paziente potrebbe inviare dal suo smartphone una foto, in modo tale da ottenere un'immediata diagnosi, senza ricorrere a canali medici destinati ai casi più urgenti [1].

Tecniche di Machine Learning comunemente usate in campo medico

Di seguito verranno illustrate le tecniche di ML più usate in letteratura medica. Si sottolinea che in questa sede avverrà una trattazione maggiormente orientata al campo medico, non approfondendo gli aspetti tecnici.

Support Vector Machine

Le SVM vengono utilizzate principalmente per classificare i soggetti all'interno dei due gruppi, aventi rispettivamente come "etichetta" $Y = 1$ e $Y = -1$ [5]. Questi raggruppamenti sono definiti da un *confine decisionale (decision boundary)* definito dai dati X in ingresso:

$$a_i = \sum_{j=1}^p w_j X_{ij} + b$$

L'obiettivo dell'addestramento della SVM è quello di rinvenire il parametro w ottimale, in modo tale che la classificazione risulti il più precisa possibile (Figura 1). Una delle proprietà più importanti delle SVM consiste nel fatto che la determinazione dei parametri è un problema di ottimizzazione convessa, dunque la soluzione è sempre un ottimo globale.

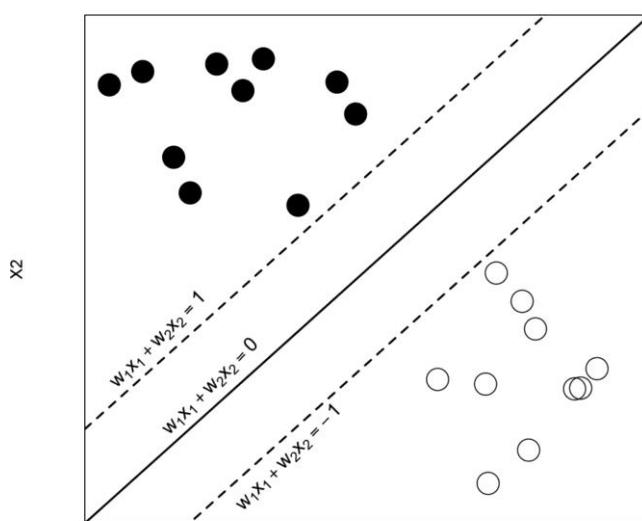


Figura 1: un esempio di come funziona una Support Vector Machine [5]

Convolutional Neural Network

L'accrescimento delle capacità computazionali dei dispositivi moderni ha permesso al Deep Learning di divenire uno dei campi di ricerca più gettonati all'interno di varie discipline scientifiche: in questo senso, la medicina non ha conosciuto eccezioni [5][6]. Grazie a queste tecniche di apprendimento "profonde" (cosiddette a causa della numerosa presenza di strati, in grado di astrarre schematicità molto complesse), si può effettuare un'analisi dettagliata di immagini mediche, come le scansioni a raggi X, sfruttando l'abilità di una rete neurale di gestire dati voluminosi e estremamente complessi, come appunto le immagini, in modo efficiente [5]. Nel corso degli anni, le Convolutional Neural Network hanno ottenuto un'enorme popolarità, specie nel mondo medico: basti pensare che dal 2014 in poi, questo particolare tipo di reti neurali ha soppiantato metodologie quali le Recurrent Neural Network e le Deep Belief Neural Network [5] (Figura 2).

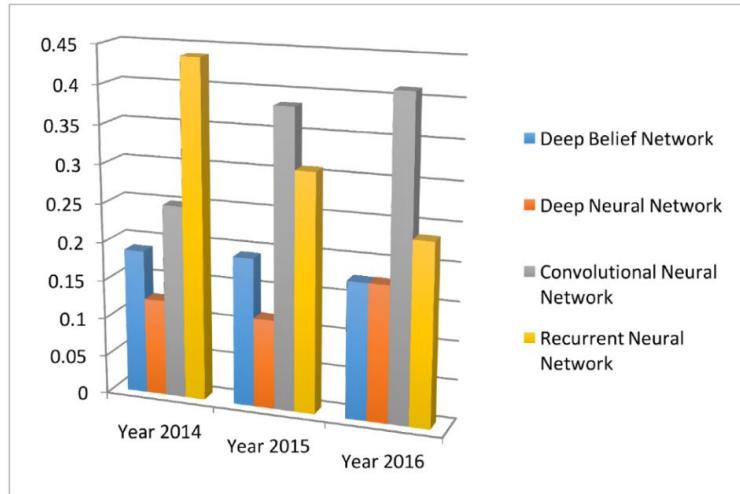


Figura 2: popolarità degli algoritmi di Deep Learning nel campo medico [5]

Una CNN si basa sull'utilizzo di un'operazione detta di *convoluzione*, in grado di tenere traccia dei vari cambiamenti di un dato multidimensionale, avendo la posizione di un misuratore all'interno di uno spazio [7]. Considerando un'immagine bidimensionale I , e un kernel K (un array multidimensionale riportante i parametri appresi dall'algoritmo):

$$S(i, j) = (I * K)(i, j) = \sum_m \sum_n I(m, n)K(i - m, j - n)$$

In figura 3 viene riportata una struttura a grandi linee di una CNN, mentre nella figura 4 è possibile notare ciò che viene prodotto tramite l'applicazione dell'operazione di convoluzione.

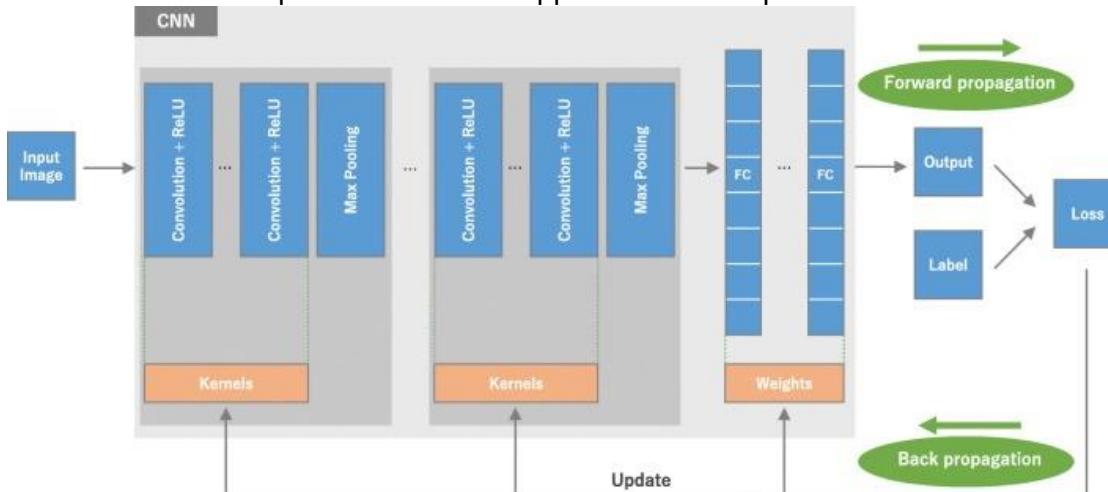


Figura 3: bozza di una CNN [8]

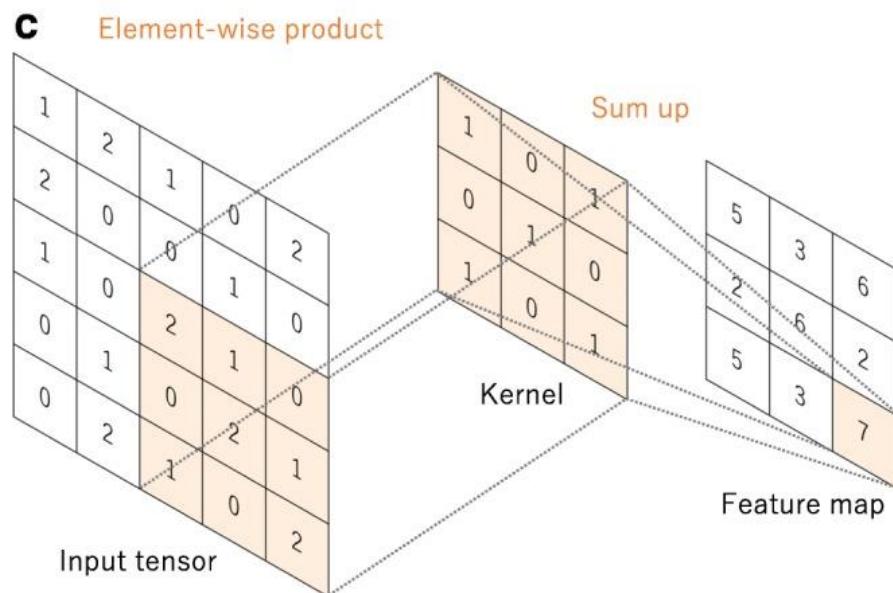
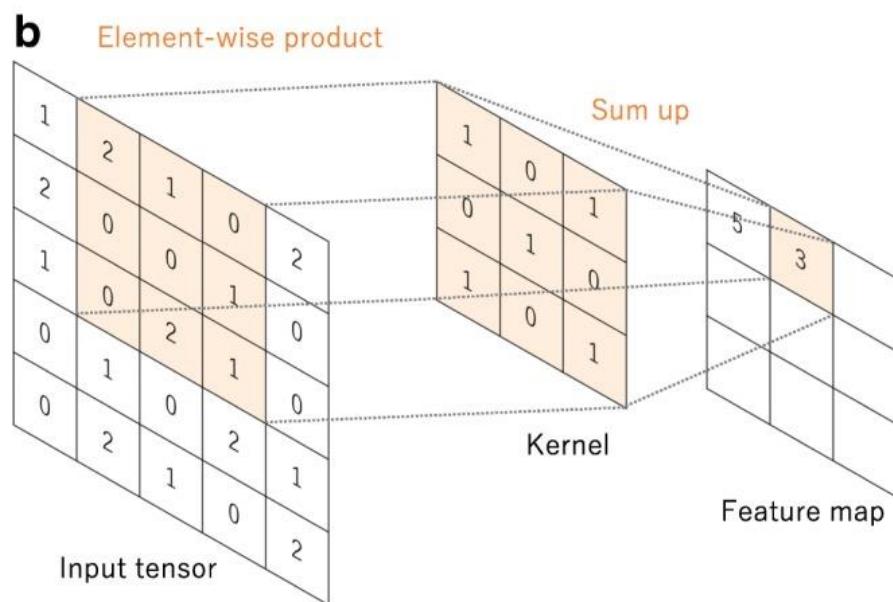
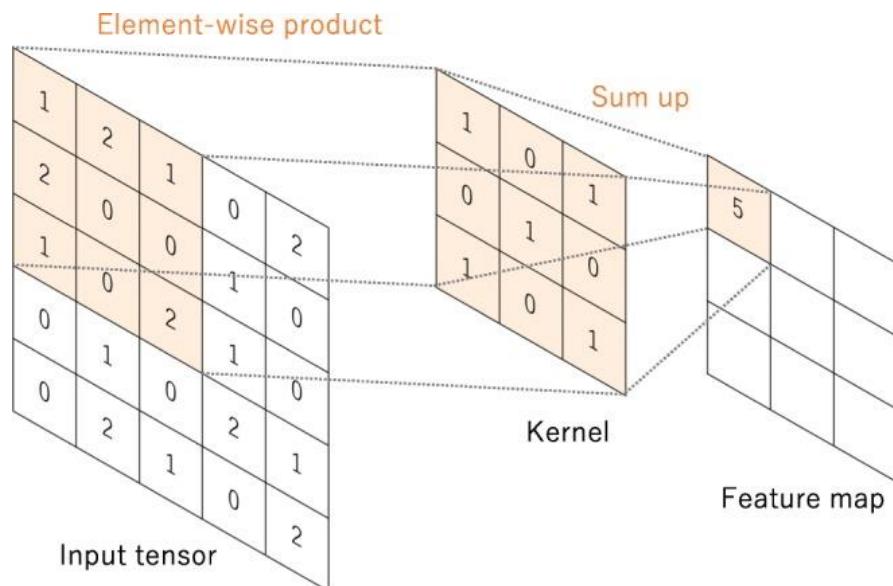


Figura 4: risultato di alcune operazioni di convoluzione[8]

Per maggiori dettagli, consultare l'articolo sulla [Smart Agriculture](#), pubblicato precedentemente su questo portale.

Random Forest

La Random Forest è una tecnica che contempla l'utilizzo di più regressori, strutturati "ad albero" (i.e.: gli *alberi di decisione*). Ogni albero esprime un proprio candidato attraverso un algoritmo di classificazione: successivamente, i voti di tutti gli alberi vengono sottoposti a una media.

$$\hat{f} = \frac{1}{B} \sum_{b=1}^B f_b(x')$$

Nell'equazione il termine B sta ad identificare i singoli *bagging*, ovvero gli addestramenti degli alberi di decisione su differenti istanze del dataset [9]. In campo medico, questo tipo di tecnica può essere utilizzata, ad esempio, per discriminare caratteri fenotipici di un organismo [10], oppure per classificare i dati clinici di un paziente in modo tale da poter fornire una diagnosi accurata [11].

Limiti dell'AI in campo medico

Disponibilità di dati di qualità

Uno dei temi centrali nella costruzione di un modello ML è quello di poter attingere a un dataset rappresentativo di tutti i possibili soggetti, in modo che sia il più diversificato possibile [1]. L'ideale sarebbe riuscire ad addestrare algoritmi utilizzando dati molto simili, se non identici, a quelli riportati all'interno delle cartelle cliniche elettroniche [1]: purtroppo molte volte si avrà a che fare con dataset piccoli, raccolti da piccoli centri clinici, talvolta di qualità scarsa (riportanti del *rumore*, ossia delle irregolarità generate da dei dati erronei).

Privacy

Come detto, avere dataset costituiti da cartelle cliniche correttamente compilate sarebbe l'ideale. Tuttavia, questi dati agli occhi della legislazione vigente sono considerati *sensibili*, dunque di difficile reperibilità, rendendo più difficile la delineazione di modelli di ML. Una soluzione naturale potrebbe essere quella di consegnare i dati clinici al paziente stesso, che deciderà poi che uso farne [1].

Apprendimento da pratiche passate sbagliate

Tutte le attività umane sono sottoposte involontariamente a dei bias cognitivi: alcuni degli aspetti da considerare nello sviluppare un sistema ML è quello di comprendere quanto questi bias, rappresentati dai dati, influenzino il modello finale [12] e quali strumenti porre in campo per far fronte a questa problematica [13].

Esperienza nella valutazione finale

Similmente a quanto avviene nei sistemi sanitari, l'applicazione di tecniche di ML richiede una sofisticata struttura regolatoria, in grado di assicurare il giusto impiego degli algoritmi nel campo medico [1]. I medici e pazienti devono capire i limiti di questi strumenti, come ad esempio l'impossibilità di una determinata struttura di poter generalizzare su un altro tipo di problematica [13]. L'affidarsi ciecamente ai modelli ML può condurre a decisioni erronee: ad esempio un medico

potrebbe abbassare la guardia nel caso in cui l'algoritmo restituisca un risultato sbagliato, al di sotto di una certa soglia d'allarme [1].

Cooperazione interdisciplinare

Team di informatici, biologi e medici devono collaborare in modo da poter costruire modelli utilizzabili nei loro rispettivi settori. Una mancata comunicazione può condurre a risultati inutilizzabili da parte dei medici [1]. Spesso le pubblicazioni scientifiche vengono pubblicate online sottoforma di prestampa (i.e.: *preprint*) su portali quali [arXiv](#) e [bioRxiv](#), senza considerare la moltitudine di manoscritti informatici che non vengono pubblicati sui giornali scientifici tradizionali ma bensì all'interno di conferenze quali [NeurIPS](#) e [ICML](#) [1].

Tecniche di AI applicate al COVID-19

Dall'inizio della pandemia del virus SARS-CoV-2, il mondo scientifico ha posto la sua attenzione nel porre in campo metodi in grado di poter contrastare l'accrescimento dei contagi: basti pensare che soltanto nel mese di marzo, sono stati pubblicati 24000 prestampati sui portali arXiv e bioRxiv concernenti l'utilizzo di tecniche di AI con il compito di individuare pazienti affetti da COVID-19 [15]. Molti autori hanno realizzato lavori in cui le CNN vengono usate per discriminare i malati COVID da altri con patologie più comuni, utilizzando dataset costituiti da scansioni a raggi X di pazienti affetti da polmoniti (Figura 5) [16][17].

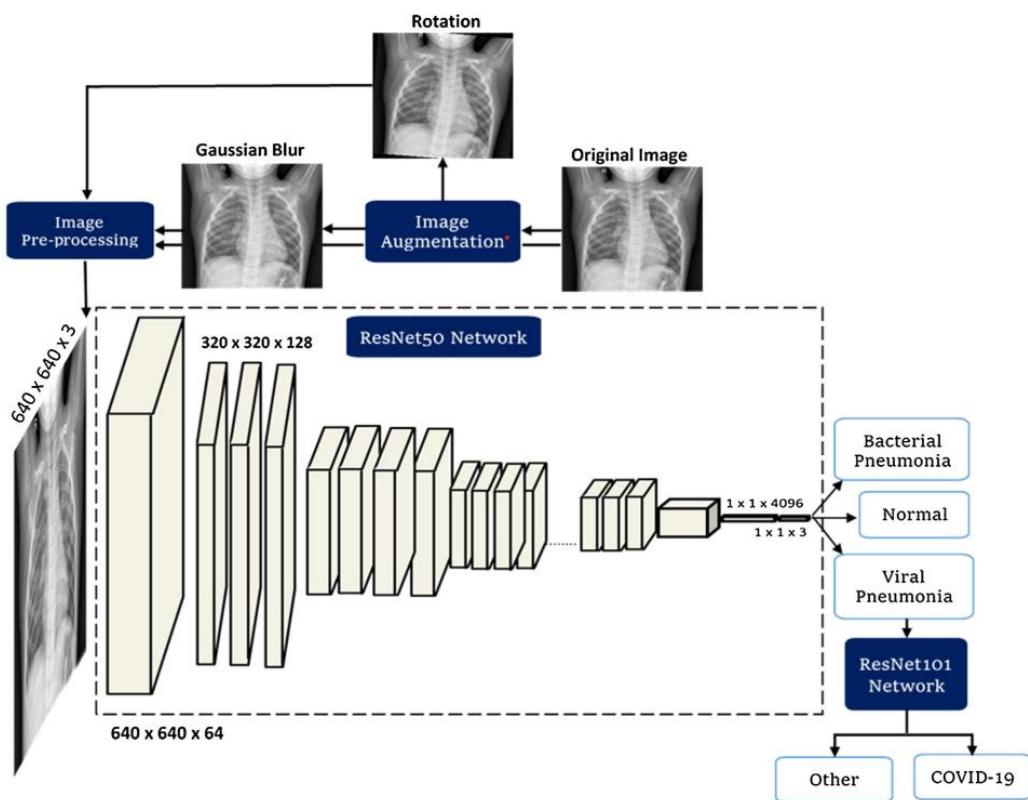


Figura 5: struttura di una CNN per l'individuazione di pazienti affetti da SARS-CoV-2[16]

Inoltre, sono stati proposti diversi metodi in real-time per l'individuazione immediata della malattia: l'utilizzo degli smart-watch, ad esempio, permette il monitoraggio di diversi parametri fisici in grado di indicare se il soggetto abbia contratto il virus o meno [18].

Conclusioni

Marco, un paziente di 49 anni, sente un dolore alla spalla e, nonostante ciò, decide di non richiedere assistenza medica. Qualche mese dopo, decide di farsi visitare da un dottore che gli diagnostica una cheratosi seborroica (un'escrescenza cutanea simile a un grosso neo). Successivamente, Marco si sottopone a una colonoscopia e un'infermiera nota una macchia scura sulla sua spalla. Marco decide di farsi visitare da un dermatologo, il quale ottiene un campione dell'escrescenza: le analisi effettuate evidenziano una lesione benigna del pigmento. Il dermatologo, però, non si fida e decide di effettuare una seconda analisi: stavolta la diagnosi parla di un melanoma invasivo. Dunque, un oncologo sottopone Marco a un trattamento chemioterapico ma, nel frattempo, un amico del dottore chiede al povero paziente come mai non si sia ancora sottoposto a immunoterapia [1].

Se Marco avesse avuto accesso alle più moderne tecnologie di ML, avrebbe potuto semplicemente fare una foto alla spalla tramite il suo smartphone per poi inoltrare l'immagine a un dermatologo esperto tramite un'app dedicata. Successivamente a una biopsia della lesione, raccomandata dal dermatologo, si sarebbe giunti alla diagnosi di melanoma al primo stadio: a quel punto, il dermatologo avrebbe potuto recidere la lesione [1].

L'applicazione dell'intelligenza artificiale in medicina potrà far risparmiare tempo, impiegare al meglio il know-how degli specialisti, permettere di giungere a diagnosi maggiormente accurate e, più in generale, migliorare la vita dei pazienti e snellire il lavoro degli operatori del settore.

- Pros:
 - Diagnosi più precise
 - Snellimento dei procedimenti medici e burocratici
 - Acquisizione di una conoscenza terapeutica globale
 - Maggiore specializzazione del personale medico
- Cons:
 - Disponibilità di dati di qualità
 - Mancanza di procedure di raccolta dati condivisi
 - Componente umana fondamentale nella valutazione finale della diagnosi

References

- [1] Rajkomar, A. - Dean, J. - Kohane, I.. "Machine Learning in Medicine." *New England Journal of Medicine*, 380. 1347-1358. (2019)
- [2] Qayyum, A - Junaid Q. – Muhammad, B. - Ala A.. "Secure and Robust Machine Learning for Healthcare: A Survey." *IEEE reviews in biomedical engineering* (2020)
- [3] McGlynn, E. – McDonald, K. – Cassel C.. "Measurement is Essential for Improving Diagnosis and Reducing Diagnostic Error: A Report from the Institute of Medicine." *The Journal of the American Medical Association*, Vol. 314(23), pp. 1-2 (2015)

- [4] Das, J. - Woskie, L. - Rajbhandari, R. - Abbasi, K. - Jha, A.. "Rethinking assumptions about delivery of healthcare: implications for universal health coverage." *The British Medical Journal (Clinical research ed.)*, Vol. 361 (2018)
- [5] Jiang, F. - Jiang, Y. - Zhi, H. - Dong, Y. - Li, H. - Ma, S. - Wang, Y. - Dong, Q. - Shen, H. - Wang, Y.. "Artificial Intelligence in Healthcare: Past, Present and Future." *Stroke and Vascular Neurology* Vol. 2 (2017)
- [6] Mori, J. - Kaji S. - Kawai H. - Kida S. - Tsubokura M. - Fukatsu, M. – Harada, K. - Noji H. - Ikezoe T. - Maeda T. - Matsuda A.. "Assessment of dysplasia in bone marrow smear with convolutional neural network." *Scientific Reports* Vol.10 (2020)
- [7] Goodfellow, I. - Bengio, Y. - Courville, A.. "Deep Learning." MIT Press (2016)
- [8] Yamashita, R. - Nishio, M. - Do, R., - Togashi, K.. "Convolutional neural networks: an overview and application in radiology." *Insights into imaging*, Vol. 9(4), pp. 611–629 (2018)
- [9] Breiman, L.. "Random Forests." *Machine Learning* Vol.45(1), pp. 5 – 32 (2001)
- [10] Chen, T. - Cao, Y. - Zhang, Y. - Liu, J. - Bao, Y. - Wang, C. - Jia, W. - Zhao, A.. "Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection." *Evidence-based complementary and alternative medicine : eCAM* (2013)
- [11] Alam, Z. - Rahman, S. - Rahman, S.. "A Random Forest based predictor for medical data classification using feature ranking." *Informatics in Medicine Unlocked*, Vol. 15, (2019)
- [12] Gianfrancesco, MA. - Tamang, S. - Yazdany, J. - Schmajuk, G.. "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data." *Journal Of American Medical Association, Internal Medicine*, Vol.178(11), pp. 1544-1547 (2018)
- [13] Rajkomar, A. - Hardt, M. - Howell, M. D. - Corrado, G. - Chin, M. H.. "Ensuring Fairness in Machine Learning to Advance Health Equity." *Annals of internal medicine*, Vol.169(12), pp. 866–872 (2018)
- [14] Krumholz, H. M.. "Big data and new knowledge in medicine: the thinking, training, and tools needed for a learning health system." *Health affairs (Project Hope)*, Vol.33(7), pp. 1163-70 (2014)
- [15] Hao, K.. "Over 24,000 coronavirus research papers are now available in one place". MIT Technology Review, URL: <https://www.technologyreview.com/2020/03/16/905290/coronavirus-24000-research-papers-available-open-data/> (2020)
- [16] Jain, G. - Mittal, D. - Thakur, D. - Mittal, M. K.. "A deep learning approach to detect Covid-19 coronavirus with X-Ray images". *Biocybernetics and biomedical engineering*, Vol.40(4), pp. 1391-1405 (2020)
- [17] Hemdan, E.E - Shouman, M. - Karar, M.. "COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images." ArXiv, abs/2003.11055 (2020)
- Mishra, T. - Wang, M. - Metwally, A.A., Bogu G.K. - Brooks A. W. - Bahmani A. - Alavi A. - Celli A. –
- [18] Higgs, E. - Dagan-Rosenfeld O. - Fay, B. - Kirkpatrick, S. - Kellogg, R. - Gibson M. - Wang, T. - Hunting E. M. - Mamic P. - Ganz A. B. - Rolnik, B. - Li, X. - Snyder M. P.. "Pre-symptomatic detection of COVID-19 from smartwatch data." *Nature Biomedical Engineering* (2020).