

Intelligenza artificiale e medicina: La validazione clinica

Percorso ragionato di un clinico nel labirinto dell'AI

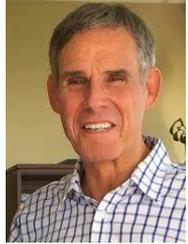


Marco Clerici
Gianfranco Gensini

«ESSENTIA» 10 dicembre 2019

Osservazioni generali...

Gli sviluppi recenti dell'innovazione tecnologica sono stati così significativi in termini sia di dimensione che di velocità che non c'è stato sempre nè il tempo nè l'opportunità di comprendere pienamente la loro sicurezza e la loro effettiva rilevanza .



NHS : The Topol Review "Preparing the healthcare workforce to deliver the digital future" February 2019

Le "istituzioni " formali e informali sono inadatte ad affrontare le sfide poste dall'AI in medicina . Vanno rinnovate per poter assicurare una governance appropriata nei confronti di queste tecnologie.



E' sempre più evidente che le decisioni cliniche sono spesso sub-ottimali e a volte fatalmente imperfette. Molte si collocano in quello spazio definito come "arte" della medicina : nel senso che il processo decisionale è basato su conoscenze incomplete o inconsistenti , su abilità, training e esperienza molto diverse. Non è sorprendente che il risultato sia una inaccettabile variabilità di qualità di cura non dipendente da fattori legati al paziente o dal contesto clinico.



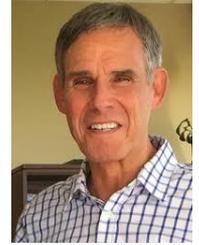
I medici sono inondati da nuove evidenze scientifiche praticamente ogni giorno e non è affatto semplice essere a conoscenza di tutto.

Le applicazioni A.I. possono meglio selezionare le migliori evidenze disponibili da un'enorme numero di ricerche cliniche, identificare i criteri di trasferibilità dei risultati sul singolo paziente e presentare al medico le opzioni migliori per il trattamento

AI e EBM

AI e «Desktop medicine»

Gli EHR che simboleggiano la digitalizzazione dell'assistenza medica (digital health) rappresentano ironicamente oggi i maggiori ostacoli al miglioramento delle cure (E.Topol)



Raccogliere le storie e prendere decisioni usando uno schermo e una tastiera richiede tempo e distrae l'attenzione dal paziente. Un'interfaccia che possa interagire in modo flessibile elaborando il linguaggio naturale potrebbe essere un' alternativa per raccogliere dati e fornire supporto alle decisioni.



Royal College of
General Practitioners

AI e paziente complesso / fragile

- Revisione periodica appropriatezza terapie
- Identificazione pazienti ad alto rischio effetti avversi terapie

- Aiuto nella deprescrizione ragionata
- Identificazione di “errori”

AI in supporto alla “care”

- Aiuto ai pazienti nel *decision making* condiviso
- Supporto alla *compliance* delle terapie

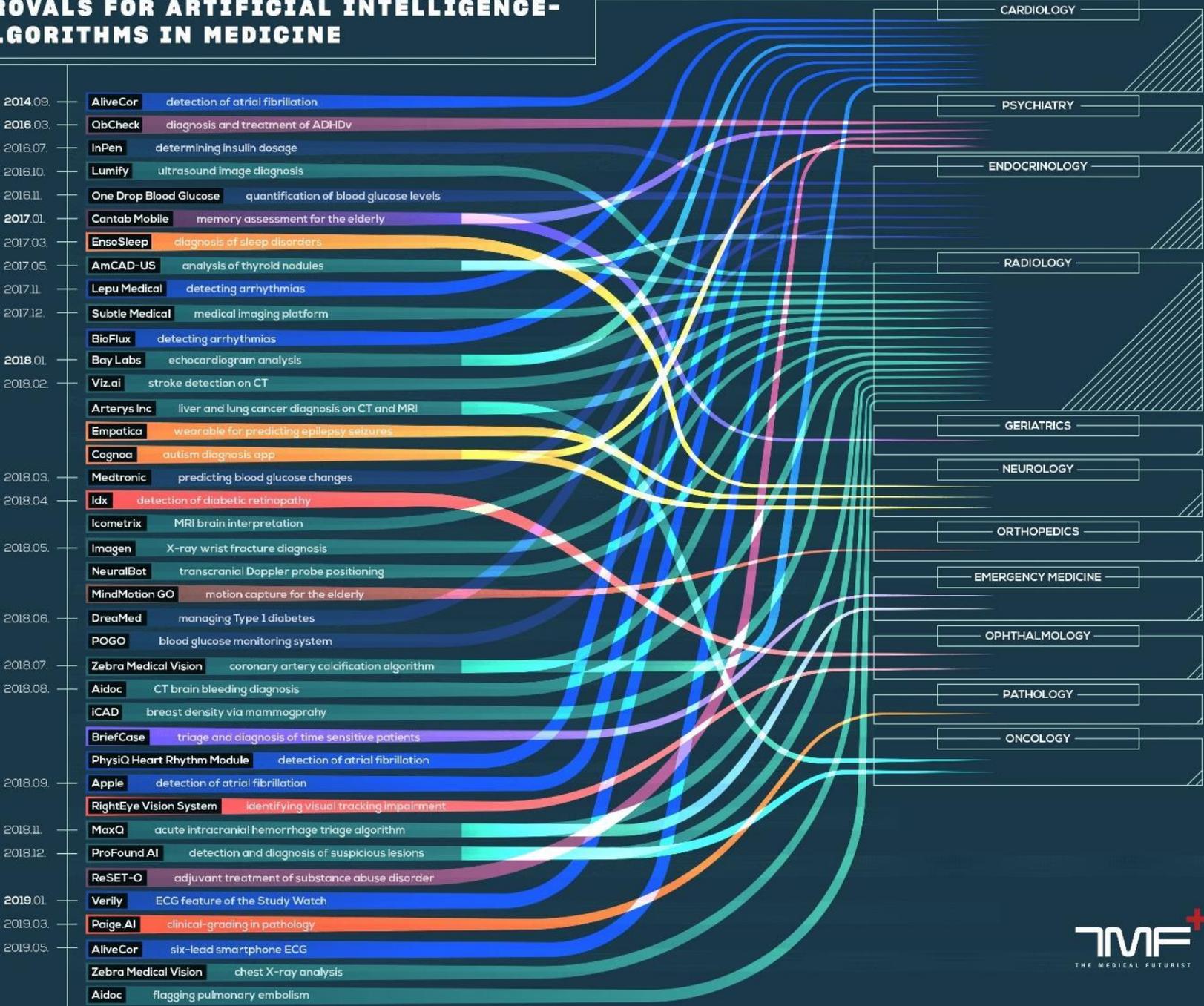
- Aiuto nell’identificazione del mix di assistenza personalizzata

*Clustering for
discovery of
disease subtypes*

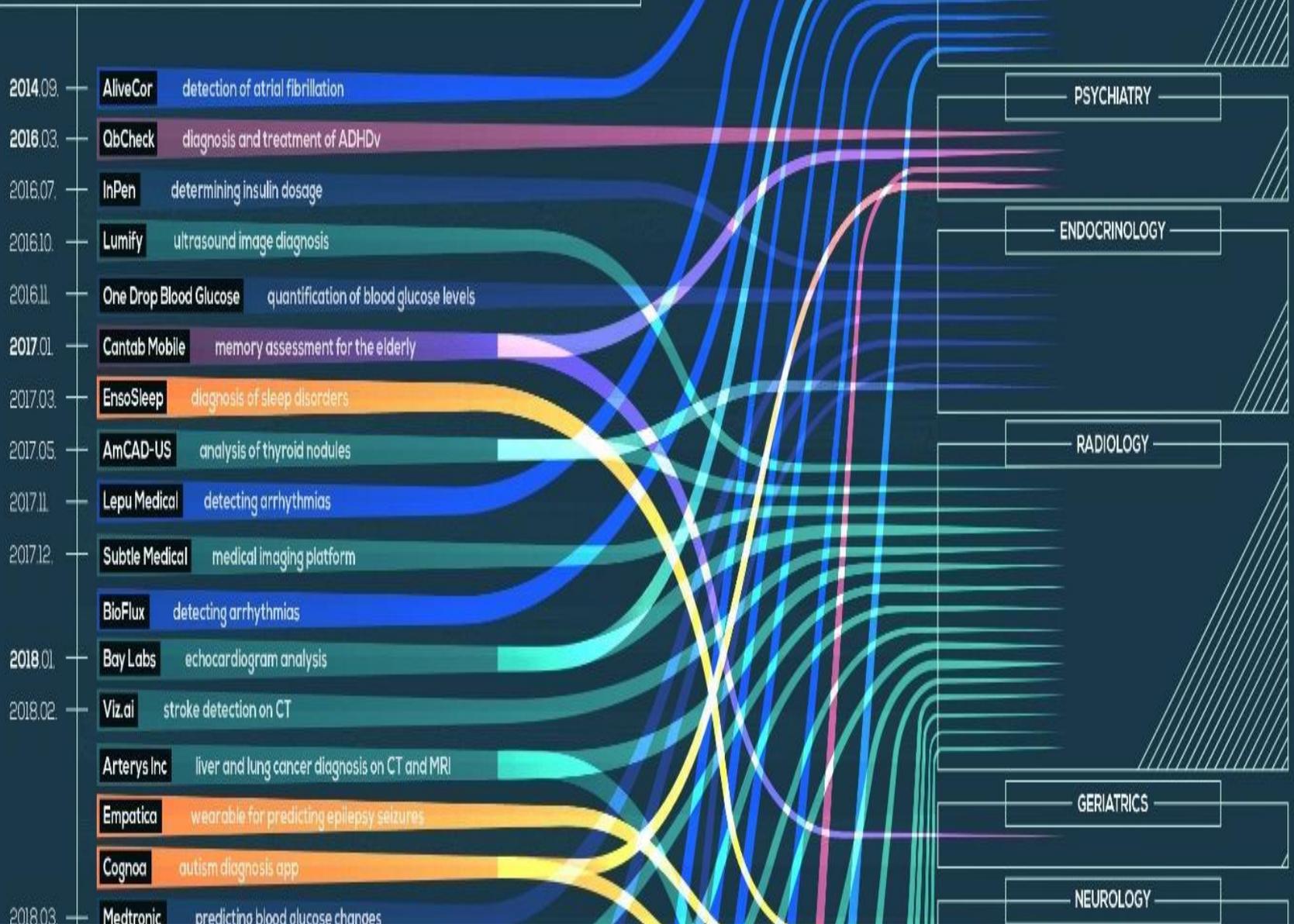


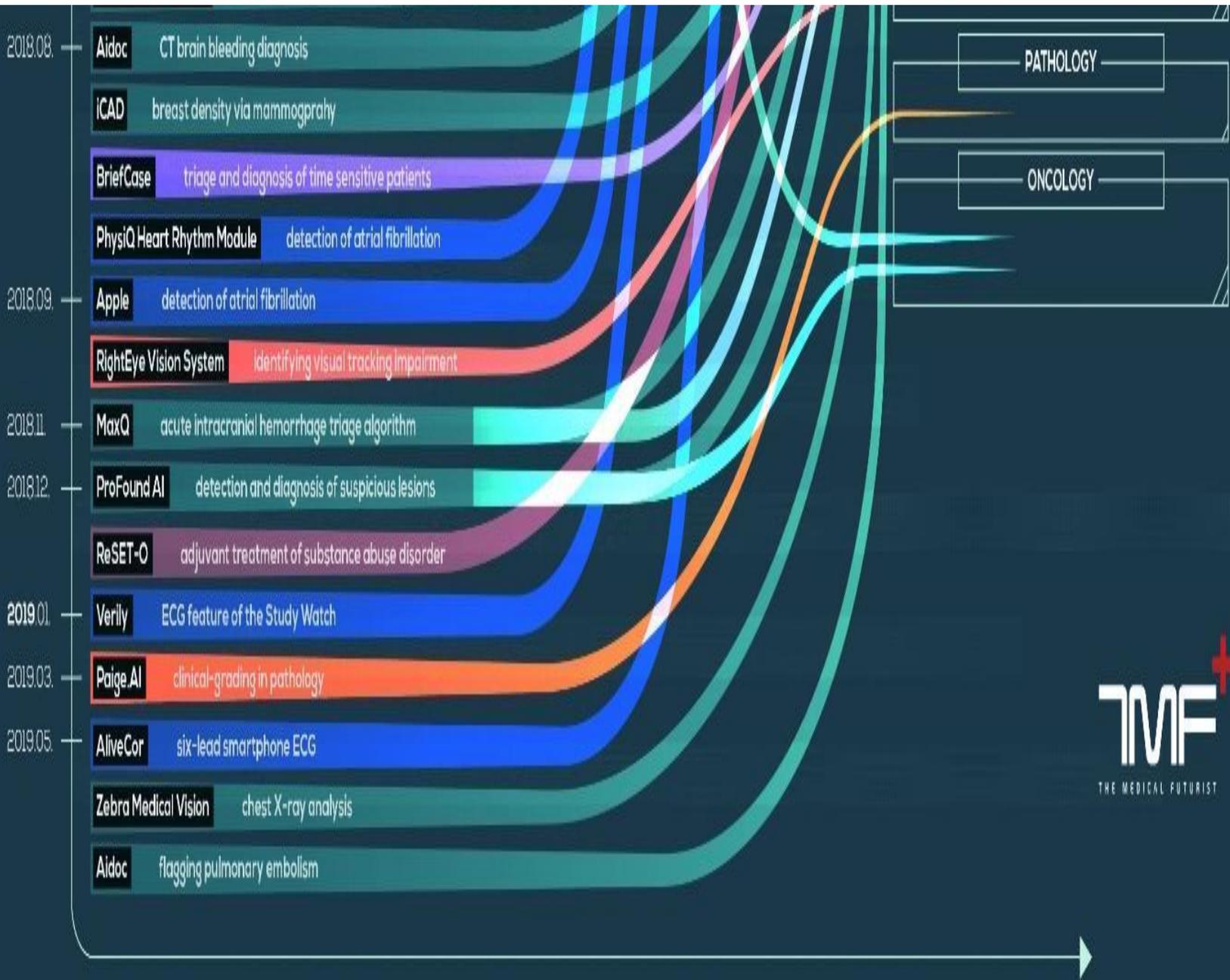
- **L'utilizzo di metodi ML per trovare nuovi sottoinsiemi di malattia utilizzabili sarà fondamentale per far progredire la medicina di precisione**

FDA APPROVALS FOR ARTIFICIAL INTELLIGENCE-BASED ALGORITHMS IN MEDICINE



FDA APPROVALS FOR ARTIFICIAL INTELLIGENCE-BASED ALGORITHMS IN MEDICINE





Validazione: connettere l' AI alla pratica clinica

La priorità deontologica : primum non nocere

1. La «regolamentazione» delle applicazioni
2. L'identificazione dei rischi
3. La rilevanza degli outcome
4. Il livello di autonomia delle «macchine»



La domanda non è tanto :“ il sistema funziona?”
ma piuttosto : “ il sistema è anche di aiuto?”

Connettere l' AI alla pratica clinica : la «regolamentazione»

Applicazioni AI: qual è il “giusto” standard di evidenza necessario?

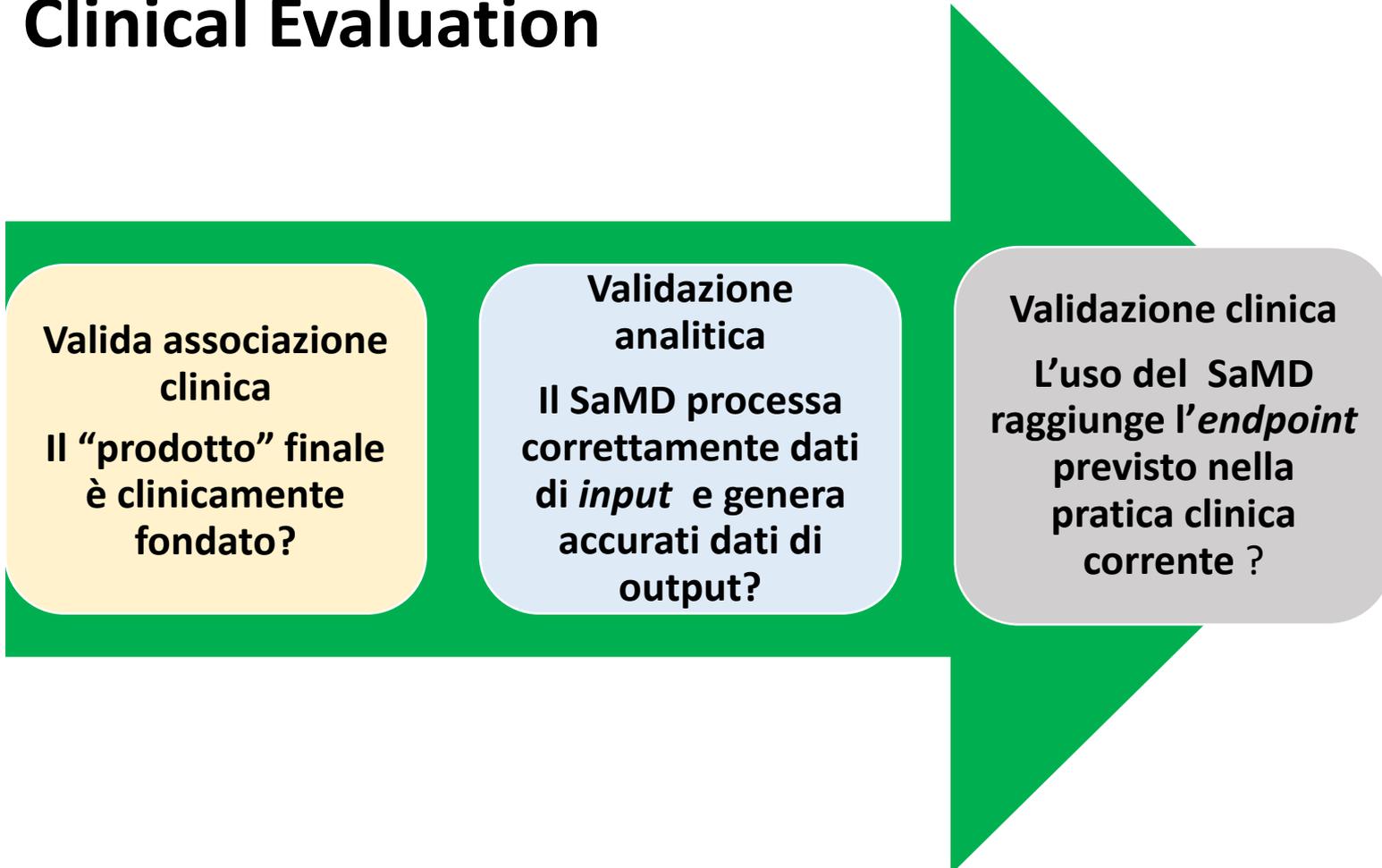
Quale il ragionevole bilanciamento tra innovazione, sicurezza ed efficacia ?

1. La verifica dell'accuratezza delle applicazioni AI di *imaging* con gli attuali standard di qualità per l'*imaging* tradizionale è probabilmente sufficiente per l'uso clinico .

2. Poiché le applicazioni dell'IA si spostano verso la previsione, la diagnosi e il trattamento, lo standard per la validazione in questi casi dovrebbe essere significativamente più alto.

3. Se un'applicazione AI influenza direttamente un outcome clinico “hard” dovrebbe essere necessario uno standard di prova rigoroso

Clinical Evaluation



Valida associazione clinica

Il “prodotto” finale è clinicamente fondato?

Validazione analitica

Il SaMD processa correttamente dati di *input* e genera accurati dati di *output*?

Validazione clinica

L’uso del SaMD raggiunge l’*endpoint* previsto nella pratica clinica corrente ?

“A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis” *October 01, 2019*



il termine “validazione” è usato in modo variabile , con alcuni autori che danno al termine il significato appropriato di validazione esterna del modello ma altri lo legano allo sviluppo del “tuning” dell’algoritmo.

Comparando la performance si nota che la validazione “interna” sovrastima l’accuratezza diagnostica rispetto alla validazione esterna. , sia del sanitari che dell’algoritmo .
Questi risultati mettono in evidenza la necessità di procedere alla validazione esterna in tutti i modelli predittivi.

Cosa importa : - la “forza” degli outcome
 - la «gravità» della condizione clinica

Condizione clinica	Impatto sulla decisione		
	Informazione sulla gestione clinica	Guida alla gestione clinica	Trattamento o diagnosi
Critica	II	III	IV (massimo rischio)
Importante	I	II	III
Non importante	I	I (minimo rischio)	II

1.FDA : Software as a Medical Device (SaMD): Clinical Evaluation – June 2017

2.FDA: Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device- February 2019

Dimostrare che le informazioni fornite dal DHT sono :

**- valide, accurate, aggiornate, riviste a intervalli regolari,
comprensibili**



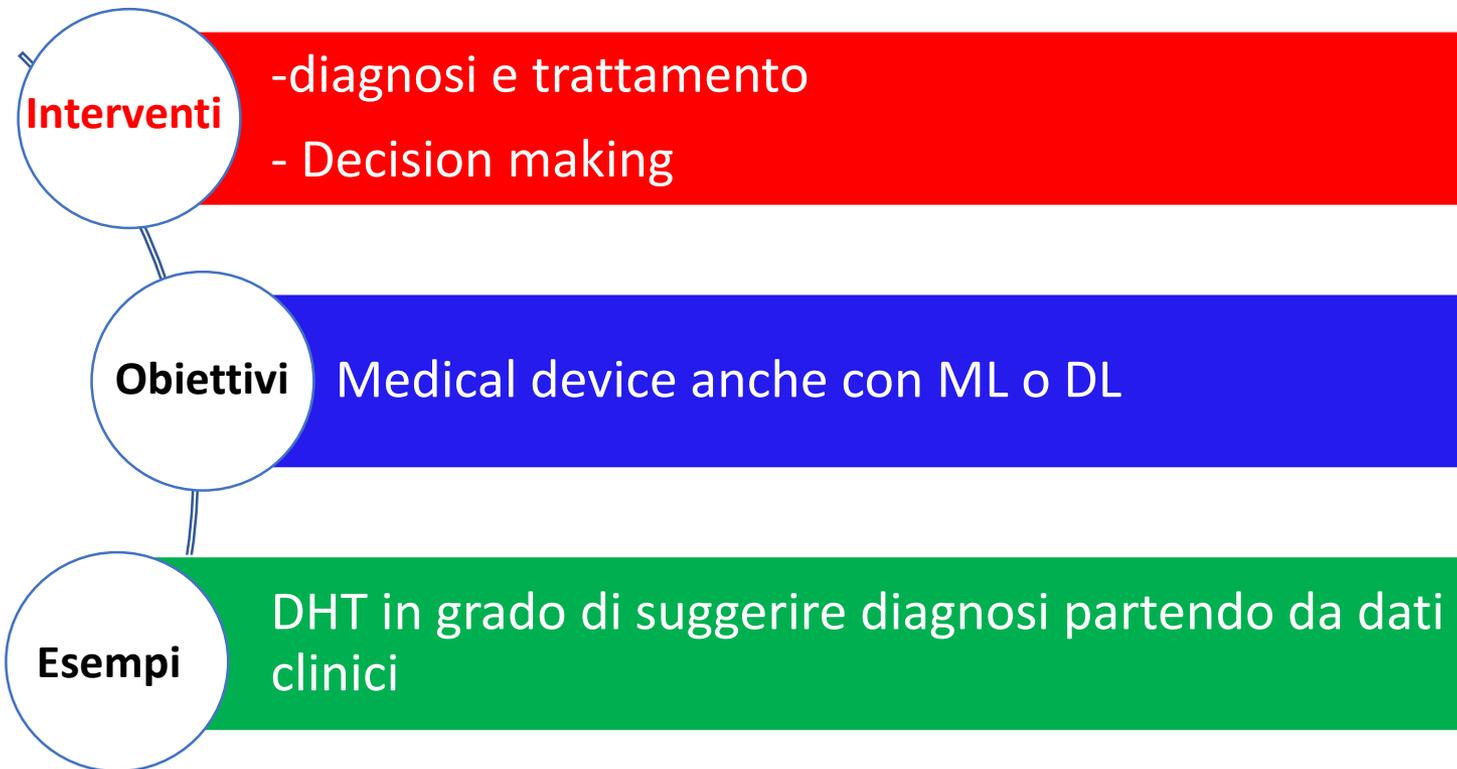
Minimum evidence standard: High quality observational studies demonstrating relevant outcomes

Best practice standard: high quality intervention study which incorporates a comparison group, showing improvements in relevant outcomes



Minimum evidence standard: high quality intervention study showing improvements in relevant outcomes

Best practice standard: High quality rct with a relevant comparator and demonstrating consistent benefit including in clinical outcomes in the target population, using validated condition-specific outcome measures.



AI E IDENTIFICAZIONE DEI RISCHI

da semplice



I primi successi dei DSS non a caso sono concentrati nelle specialità mediche **image –intensive** (radiologia [**TAC e diagnosi stroke**], oftalmologia [**screening retinopatia diabetica**], dermatologia [**diagnosi ca cute**])

**[accuratezza diagnostica comparabile a quella di medici esperti]*

a complesso



Molte azioni fondamentali in medicina, quali la previsione clinica, la diagnosi complessa [**non image-intensive**], la terapia, la riabilitazione, presentano maggiori difficoltà per le applicazioni A.I.



Dove i rischi senza «intenzionalità»?

Dov'è il «crinale scivoloso [slippery slope]?»

Applicazioni con ML (o DL) con algoritmi adattativi in diagnosi non «image intensive» e nella scelta dei percorsi di cura e/o di riabilitazione



Dove i rischi senza «intenzionalità»?



Royal College of
General Practitioners

Domande

- qual'è il miglior approccio per identificare bias o incompletezze dei dati e in continuo annullare o ridurre gli effetti di queste carenze?
- come regolare al meglio sistemi adattativi che continuano ad apprendere rispetto a sistemi in versione statica ?
- come definire e condividere il livello di autonomia delegabile a processi e sistemi automatici?
- Come definire i tipi di controllo che devono essere messi in atto nei confronti di processi e sistemi automatici ?

Dove i rischi senza intenzionalità? Se ne parla.....

Statistical bias....

Blackbox....

Social bias....

Garbage in /garbage out...

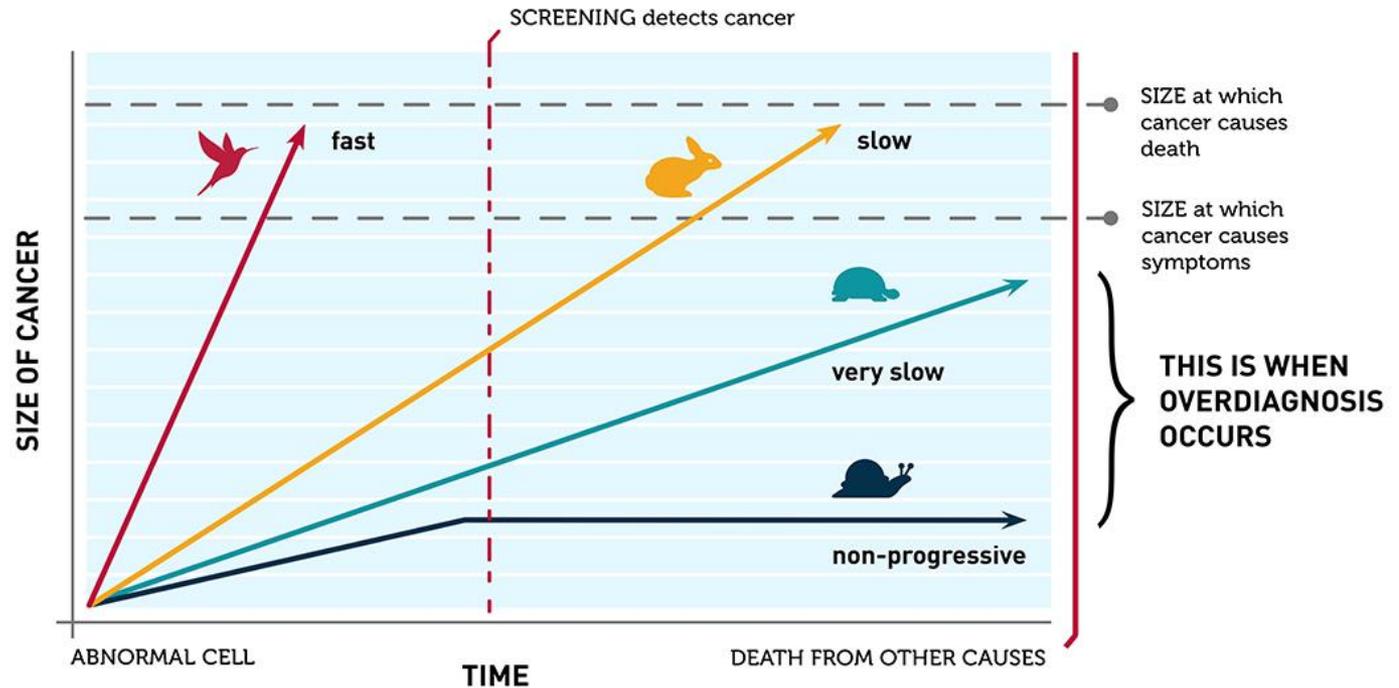
Ancora quasi non se ne parla...

La possibilità di sovradiagnosi (es da monitoraggi e screening intensivi, estensivi o comunque inappropriati)



OVERDIAGNOSIS

occurs when screen-detected cancers are either **non-growing** or so **slow-growing** that they would never cause medical problems



Adapted from a figure courtesy of
H. Gilbert Welch, Dartmouth Medical School

Screening.....



La sovradiagnosi, più che un errore che riguarda la diagnosi, è un errore che riguarda la prognosi.

Avviene quando non possiamo sapere quali conseguenze una lesione maligna individuata con un test diagnostico avrà sulla vita del paziente.

Díaz Candamio MJ «Overdiagnosis in imaging» Radiologia. 2018 Sep - Oct;60(5):362-367 .

La maggior parte dei cancro identificati con test di screening devono essere trattati. Ma alcuni NO.

Si stima, ad esempio, che il 19% dei cancro della mammella identificati con lo screening e il 20-50% dei cancro alla prostata sono a tutti gli effetti delle sovradiagnosi.

Monitoraggi intensivies rilevazione episodi F.A. «subclinica»

Nel corso di monitoraggi intensivi, c'è una significativa possibilità di generare una varietà di “segnali” di poca o nessuna rilevanza clinica.

Molta prudenza va usata nell'interpretare questi dati soprattutto prima di considerare l'implementazione di questi sistemi nella pratica quotidiana.



La sovradiagnosi è un problema crescente, che può ulteriormente aggravarsi come risultato di miglioramenti nella sensibilità dei metodi di identificazione

AI E RILEVANZA DEGLI OUTCOME



responsabilità



anche se i risultati con A.I saranno migliori dell' "usual care" non saranno mai esenti da errori che potrebbero generare conseguenze negative rilevanti

allocazione



Non è ancora chiaro se le molteplici possibili applicazioni , con i fabbisogni di qualità dei dati di input, di conservazione degli stessi e del loro continuo aggiornamento finiranno con il ridurre significativamente i costi.

Molti strumenti A.I. potrebbero semplicemente, a parità di risultati, trasferire costi

Rodgers H «[Robot assisted training for the upper limb after stroke \(RATULS\): a multicentre randomised controlled trial.](#) *Lancet.* 2019 Jul 6;394(10192):51-62

RATULS il primo rct multicentrico con adeguata «potenza statistica» per confrontare una riabilitazione robot-assistita vs un programma identico (EBM based- EULT) vs «usual care» [770 pazienti post ictus]

La riabilitazione robot-assistita non ha migliorato la funzione motoria dell'arto sup vs gli altri 2 bracci di intervento e quindi non se ne supporta l'uso clinico

Analisi costo-efficacia:

robot-assisted training £5387, EULT £4451, usual care cost £3785 per partecipante

Cosa importa : la rilevanza degli outcome

Rilevanza degli outcome in ordine decrescente:

- Mortalità generale (o sopravvivenza per determinato tempo)
- Mortalità specifica per malattia
- Eventi clinici maggiori: es stroke, IMA ecc.
- Qualità di vita (attentamente definita, NON soddisfazione)
- Outcomes surrogati: es colesterolemia, P.A. ecc.



Le applicazioni AI devono dimostrarsi utili ai pazienti e al sistema sanitario. Se i vantaggi sono insignificanti, la proliferazione non sistematica e non controllata di strumenti AI con valore clinico aggiunto minimo aumenterà solo le consultazioni, le sovradiagnosi e i sovratrattamenti e , in definitiva, i costi sanitari complessivi, nonché la sfiducia nei confronti della tecnologia.

L. A. Celi :*"An awakening in medicine: the partnership of humanity and intelligent machines"*

www.thelancet.com/digital-health Vol 1 October 2019

AI E RILEVANZA DEGLI OUTCOME

Non ci sono dati che suggeriscono che cambiare la scala di una previsione :

- **il tuo fumare ti fa rientrare in una fascia ad alto rischio di cancro**
- **c'è una probabilità del 27,6% che il tuo fumare ti causerà un cancro al polmone**

avrà maggiore successo nel modificare i comportamenti

Il problema più pressante è quello di riuscire a cambiare i comportamenti di milioni di pazienti e di migliaia di clinici.

Il comportamento dei medici è responsabile dell'80% dei costi del sistema.

Il comportamento di milioni di pazienti influenza per più del 50% gli outcome relativi alle patologie croniche.

Un focus limitato a dati e analisi può distrarre il sistema sanitario da quello che è necessario conseguire : una trasformazione di senso della "care".

AI E RILEVANZA DEGLI OUTCOME

[*Babylon Healthcheck*]



**AI e
rischio
equità?**

I pazienti utilizzatori di BGPaH sono più sani di quelli delle altre practice, più giovani, più ricchi . Sono anche maggiori utilizzatori di NHS 111

Più della metà hanno utilizzato il symptom checker online. Intervistati non pensano che il symptom checker abbia rimpiazzato la necessità di una consultazione con il GP

Si può dedurre che il servizio non è stato scelto dai pazienti anziani o dalle persone con bisogni sanitari più complessi



AI E I SISTEMI DI SUPPORTO DECISIONALE



**Chi
decide
cosa è
«futile»?**

Devono “funzionare” in un ambiente dove l’incertezza è strutturale. Per molte decisioni cliniche non c’è una risposta univoca «giusta»

Trascurare elementi poco o per niente “datificabili” (aspetti culturali, sociali o psicologici) potrebbe indurre i medici a sottostimare la funzione di questi elementi essenziali per un’accurata «care»

« Caring by design? »

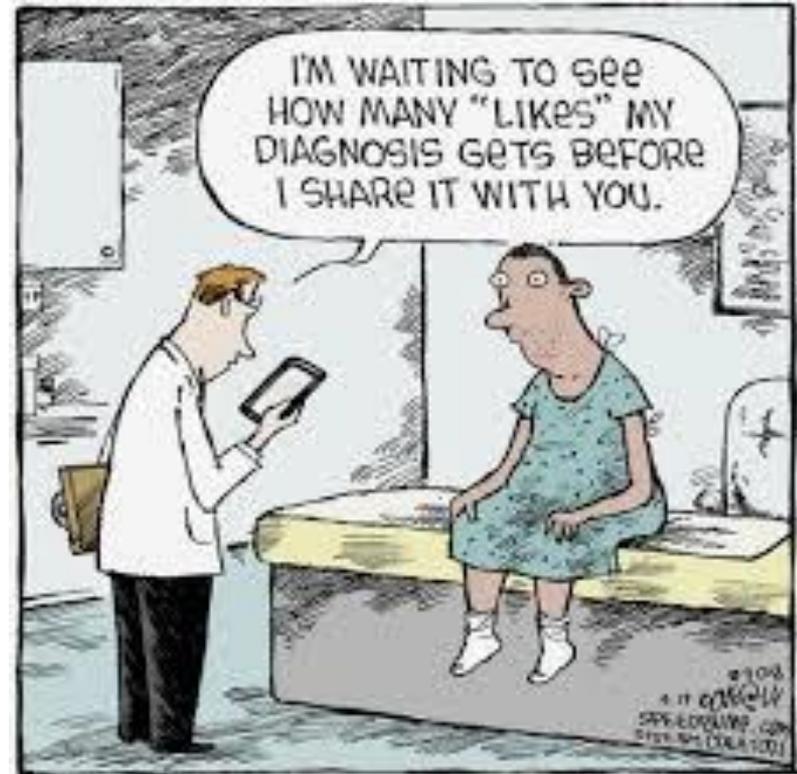
Domande:

- tutti gli attori coinvolti ne accetteranno l'introduzione ?
- come li utilizzeranno nella pratica quotidiana?
- fino a che punto alcuni ruoli professionali saranno ridisegnati ?
- come si modificheranno le relazioni tra medici?
- e il rapporto medico-paziente?
- quali effetti comporterà l'impiego del ML in termini di formazione medica?
- su quali obiettivi clinici sarà misurata l'efficacia?

Criticità dei sistemi «oracolari» in medicina

Rischio che nel tempo i medici possano sviluppare un ingiustificato ed eccessivo affidamento nelle capacità dell'automazione (over-reliance)

Rischio di sviluppare una vera e propria dipendenza (overdependence) da questi sistemi che, nel lungo periodo, potrebbe condurre alla dequalificazione (deskilling)



L'A.I. nella pratica clinica dovrebbe evolvere come “un'avventura ibrida” con i clinici che sanno cosa aspettarsi, come lavorare con uno strumento di supporto fondamentale e molto sofisticato

Lavorando assieme, umani e macchine possono affrontare molte delle debolezze decisionali intrinseche nella pratica clinica corrente. I risultati della ricerca clinica possono essere potentemente migliorati da metodi computazionali in grado di vagliare le correlazioni in grandi quantità di dati longitudinali generati dai pazienti e dai sistemi sanitari.

Ma la guida di questi metodi richiede anche definizioni più precise di alcuni fondamentali principi in medicina: cos'è normale, quali anomalie richiedono un intervento clinico, quali outcome stiamo cercando di avere, e quali costi sono accettabili per ottenerli.