

## LA CHIMICA ARTIFICIALE

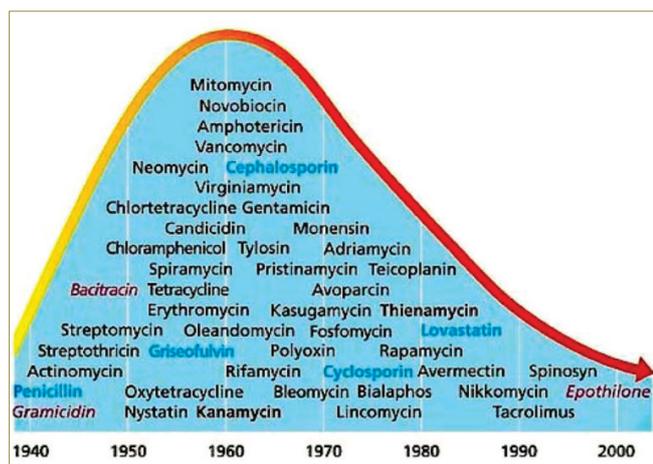


Fig. 1 - Tratta da B. Gonzalez-Zorn, J.A. Excudero, *Int. Microbiol.*, 2012, 15, 101; DOI: [10.2436/20.1501.01.163](https://doi.org/10.2436/20.1501.01.163)

Uno dei lavori che mi ha colpito di più recentemente è stato quello pubblicato su *Cell Press* [J.M. Stokes *et al.*, *Cell.*, February 2020, **180**(4), 688-702. e13; doi:[10.1016/j.cell.2020.01.021](https://doi.org/10.1016/j.cell.2020.01.021)] e dedicato alla recente scoperta dell'halicina, un nuovo antibiotico, del tutto originale e, soprattutto, scoperto in modo un po' diverso dal solito ma che potrebbe diventare in futuro un procedimento comune.

Come sapete in anni recenti la ricerca e la sintesi di nuovi antibiotici ha segnato il passo a causa dei costi elevati pur in presenza della necessità di far fronte alla crisi degli antibiotici tradizionali ed alla crescita delle resistenze soprattutto in ambito ospedaliero; è una situazione grave che costa decine di migliaia di vittime all'anno.

Come si vede dal grafico di Fig. 1 c'è stato un vero e proprio "picco degli antibiotici" negli anni passati dovuto ai costi crescenti e ai tempi lunghi della ricerca, mentre l'uso incontrollato e i meccanismi di resistenza facevano il resto. Una delle ultime scoperte importanti è stata nel 2015 e l'abbiamo presentata sul blog, [la teixobactina](#).

Ma il problema rimane.

La halicina deve il suo nome al modo in cui è stata scoperta che ha a che fare con HAL9000, il computer di 2001 *Odissea nello spazio*, (Heuristic ALgorithmic)

l'intelligenza artificiale per antonomasia (che però, ricordiamolo sempre, alla fine si svela come un ribelle e un pericolo per gli uomini che lo stavano usando, ma questa è un'altra storia...).

L'halicina, SU3327, sintetizzata nel 2015 è un potente inibitore della kinasi c-Jun N-terminale [S.K. De *et al.*, *Jour. Med. Chem.*, 2009, **52**(7), 1943; Jang *et al.*, *Redox Biol.*, 2015, **6**, 552] ed è strutturalmente diversa dagli antibiotici convenzionali. In origine era stata sintetizzata come antidiabetico, ma con scarsi risultati. Il lavoro che abbiamo citato ha dimostrato che l'halicina, individuata da un programma di AI (Intelligenza Artificiale) come un potenziale potente antibiotico, è effettivamente tale contro parecchi batteri (Fig. 2). Essa ha dimostrato di essere efficace contro *Clostridioides difficile*, *Acinetobacter baumannii*, *Mycobacterium tuberculosis* e anche contro batteri resistenti, con un meccanismo inusuale: aiutando a sequestrare il ferro ed interferendo con la regolazione del pH e delle proprietà di membrana. Sembra che in particolare l'halicina impedisca al batterio di mantenere l'opportuno gradiente elettrochimico attraverso la membrana cellulare, gradiente necessario a produrre ATP, che è una molecola chiave per la vita del batterio (e anche per le nostre cellule, ma gli enzimi sono diversi).

Dunque non si tratta di una nuova molecola ma di una molecola che faceva parte di quel rapidamente crescente universo fatto di ormai oltre 100 milioni di strutture già sintetizzate dall'uomo e messe da parte: una mole di conoscenze preziose, ma non ancora usate bene. L'intelligenza artificiale mostra una strada di possibile utilizzo di queste scoperte che sono, dopo tutto, lavoro umano accumulato, risultato sociale condiviso, pubblicato ma ancora non finalizzato.

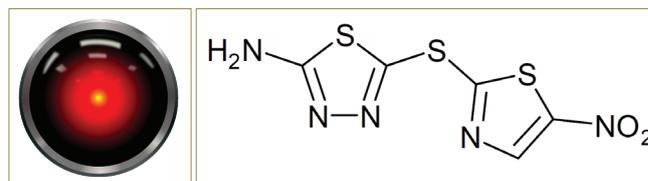


Fig. 2 - Da <https://en.wikipedia.org/wiki/Halicin>

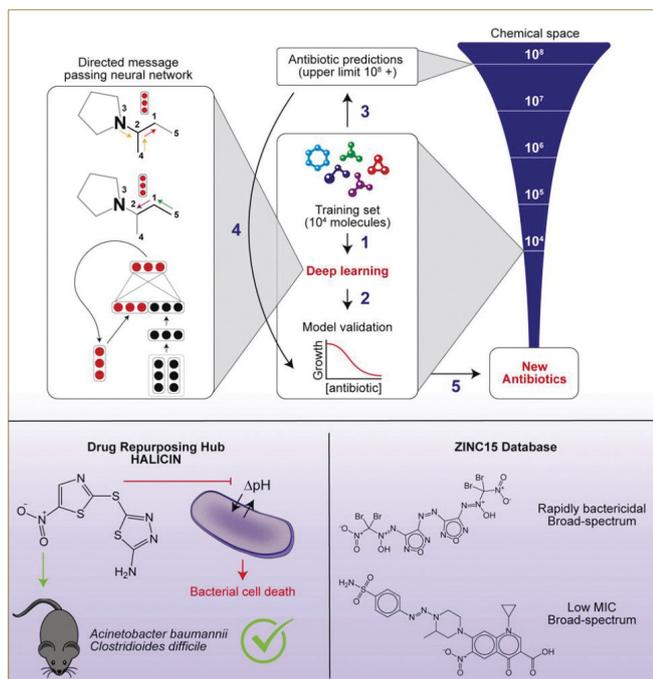


Fig. 3 - Da J.M. Stokes et al., Cell., February 2020, 180(4), 688-702. e13; doi:10.1016/j.cell.2020.01.021

Gli autori hanno “addestrato” una rete neurale in grado di prevedere quali molecole abbiano attività antibatterica. Hanno effettuato previsioni su più biblioteche chimiche e scoperto che una molecola dal Drug Repurposing Hub, halicina, che è strutturalmente divergente dagli antibiotici convenzionali, mostrava attività battericida contro un ampio spettro di agenti patogeni (Fig. 3). Inoltre, da un insieme discreto di 23 previsioni testate empiricamente da più di 107 milioni di molecole presenti nel database ZINC15, il modello ha identificato otto composti antibatterici che sono strutturalmente distanti da antibiotici noti.

Questo lavoro evidenzia l'utilità degli approcci di *deep learning* per espandere il nostro arsenale di antibiotici attraverso la scoperta di molecole antibatteriche strutturalmente distinte.

Per capire il *deep learning*, ossia l'apprendimento profondo, immaginiamo un bambino la cui prima parola è “cane”. Il bambino impara cos'è un cane (e cosa non lo è) indicando oggetti e dicendo la parola cane. Il genitore dice “Sì, quello è un cane” o “No, non è un cane”. Mentre il bambino continua a puntare agli oggetti, diventa più consapevole delle caratteristiche che tutti i cani possiedono. Ciò che il bambino fa, senza saperlo, è chiarire un'astrazione complessa (il concetto di cane) costruendo una

gerarchia in cui ogni livello di astrazione viene creato con la conoscenza che è stata acquisita dallo strato precedente della gerarchia.

A differenza del bambino, che impiegherà settimane o addirittura mesi per comprendere il concetto di “cane” e lo farà con l'aiuto del genitore (quello che viene definito apprendimento supervisionato), un'applicazione che utilizza algoritmi di *deep learning* può mostrare e ordinare milioni di immagini, identificando con precisione quali immagini contengono i set di dati, in pochi minuti pur non avendo avuto alcun tipo di indirizzamento sulla correttezza o meno dell'identificazione di determinate immagini nel corso del training.

Il *deep learning* è un sistema che sfrutta una classe di algoritmi di apprendimento automatico che usano vari livelli a cascata per svolgere compiti di estrazione di caratteristiche e di trasformazione; sono basati sull'apprendimento non supervisionato (cioè automatico non controllato dall'operatore “padre”) di livelli gerarchici multipli di caratteristiche (e di rappresentazioni) dei dati. Le caratteristiche di più alto livello vengono derivate da quelle di livello più basso per creare una rappresentazione gerarchica. I livelli multipli di rappresentazione che corrispondono a differenti livelli di astrazione formano una *gerarchia di concetti* (Fig. 4).

L'aggettivo *profondo* significa “su più livelli” (vale a dire sul numero di strati nascosti nella rete neurale - chiamati *hidden layer* (strati nascosti): quelle “tradizionali” contengono 2-3 layer, mentre le reti neurali *deep*, ossia profonde, possono contenerne oltre 150) (Fig. 5).

Per esempio i neuroni del primo strato potrebbero imparare a riconoscere i bordi, i neuroni nel secondo strato potrebbero imparare a riconoscere forme più complesse, ad esempio triangoli o rettangoli, create dai bordi. Il terzo strato riconoscerebbe forme ancora più complesse, il quarto riconosce ulteriori dettagli

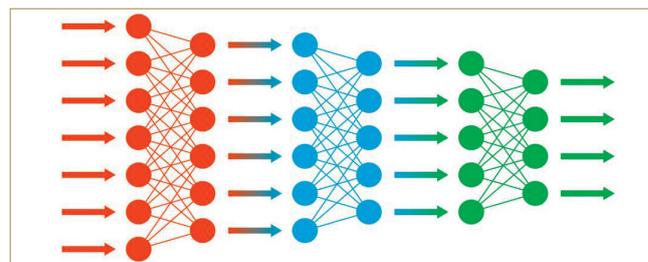


Fig. 4 - Da <https://www.ai4business.it/intelligenza-artificiale/deep-learning/deep-learning-cose/>

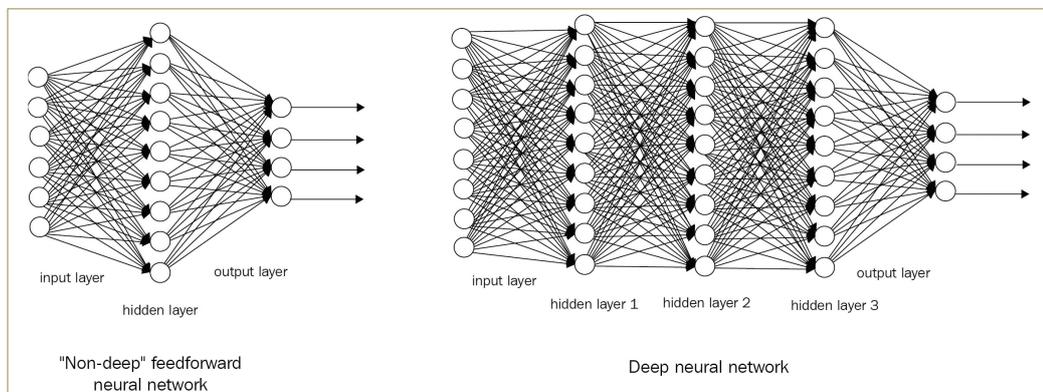


Fig. 5 - Da <https://www.kdnuggets.com/2020/02/deep-neural-networks.html>

no tanto; il ricorso a capacità computazionali via cloud può attenuare il problema perché la formazione di una rete neurale profonda richiede spesso l'elaborazione di grandi quantità di dati utilizzando cluster di GPU di fascia alta per lungo tempo.

e così via... i molteplici livelli di astrazione possono dare alle reti neurali profonde un vantaggio enorme nell'imparare a risolvere complessi problemi di riconoscimento di schemi, proprio perché ad ogni livello intermedio aggiungono informazioni e analisi utili a fornire un output affidabile.

Si capisce che quanti più livelli intermedi ci sono in una rete neurale profonda (e quindi quanto più è grande la rete) tanto più efficace è il risultato ma, di contro, la scalabilità della rete neurale è strettamente correlata ai data set, ai modelli matematici e alle risorse computazionali. Le capacità computazionali richieste sono enormi e possono rappresentare un limite; comunque i sistemi di *deep learning* migliorano le proprie prestazioni all'aumentare dei dati mentre le normali applicazioni di *machine learning* (o meglio, i cosiddetti sistemi di apprendimento superficiale, non profondo) una volta raggiunto un certo livello non sono più scalabili nemmeno aggiungendo esempi e dati di training alla rete neurale.

Nei sistemi di *deep learning* gli scienziati "etichettano" i dati (con i meta tag, i tag sono le parole aggiunte per esempio ai post del blog per indicizzarle, i meta tag una sorta di abstract del post); per esempio inserendo il meta tag "cane" all'interno delle immagini che contengono un cane, ma senza spiegare al sistema come riconoscerlo, è il sistema stesso, attraverso livelli gerarchici multipli, che intuisce cosa caratterizza un cane (le zampe, il pelo, ecc.) e quindi come riconoscerlo. Questi sistemi si basano, in sostanza, su un processo di apprendimento "tentativo-ed-errore" e, perché l'output finale sia affidabile, sono necessarie enormi quantità di dati. Ci vogliono comunque i dispositivi più potenti per il *deep learning* e costa-

Non è chiaro allora se l'impegno in termini economici sia maggiore o minore del passato, ma certo si è dimostrato che può aprire strade nuove; rimane però, e non mi stancherò di dirlo, che anche con questo nuovo mezzo di ricerca non si può affidare tutto alla gara fra batteri e antibiotici confidando nella nostra capacità di scoperta. Dobbiamo assolutamente smetterla, dopo aver fatto tanto lavoro, di lasciare al "mercato" l'uso dei risultati e, dunque, far succedere quel che poi è già successo: uso smodato e improprio dei risultati a soli scopi di profitto; dobbiamo usare razionalmente i risultati e gli antibiotici devono essere usati solo quando servono veramente; la Natura è in grado di far evolvere velocemente contro misure di resistenza nei batteri (negli insetti, nelle piante, nei virus, etc.).

Il caso coronavirus, che al tempo in cui leggerete questo articolo potrebbe essere in via di risoluzione, offre una lezione analoga; non possiamo inseguire i virus con gli antivirali o gli anticorpi monoclonali, dobbiamo metterci in un altro rapporto NON MERCANTILE con la Natura e ridurre al minimo gli *spillover*, i salti di specie, riducendo l'invasione dei sistemi naturali ancora indipendenti, le occasioni di scambio di materiale genetico fra noi e i nostri animali e i selvatici, lasciando alla natura lo spazio che le serve, non cacciando inutilmente specie protette o rare, non uccidendo o allevando o mescolando tutto in modo crescente e spropositato; se no pagheremo un prezzo altissimo.

Fraasi in italiano e alcune definizioni da <https://www.ai4business.it/intelligenza-artificiale/deep-learning/deep-learning-cose/>; e-book gratuito: <http://neuralnetworksanddeeplearning.com/chap5.html>

