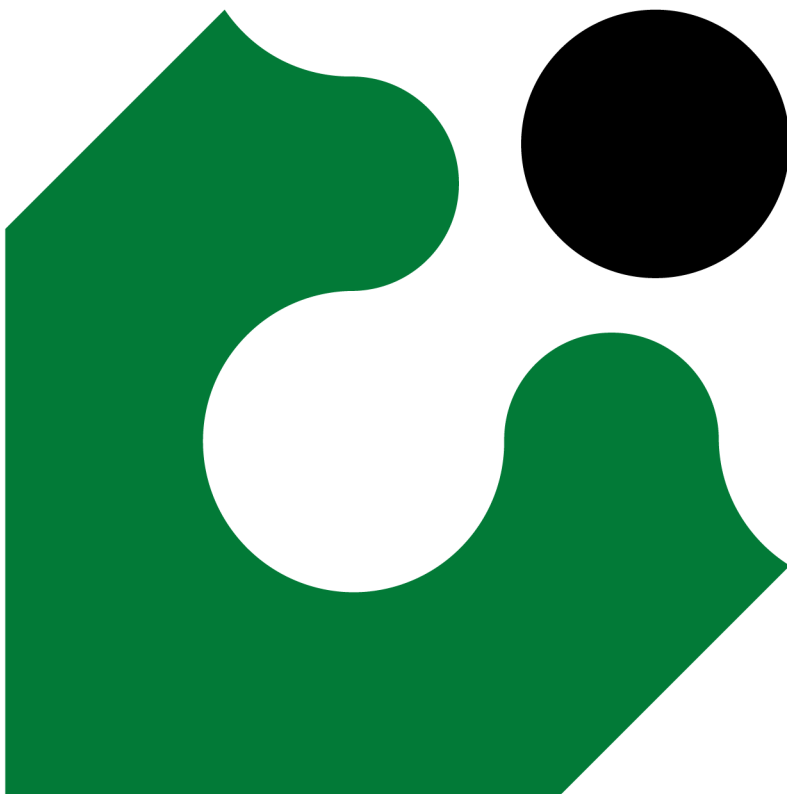




Intelligenza artificiale per le diagnosi complesse in neuro-oncologia

Dr. Gaetano Finocchiaro

Anno formativo 2020



Corso di Rivalidazione manageriale

“Lo sviluppo del manager in sanità: tra sfide e cambiamenti”

ECOLE 19001/AE

GLI AUTORI

*Dr. Gaetano Finocchiaro, Direttore UOC, Fondazione IRCCS Istituto Neurologico Besta,
gaetano.finocchiaro@istituto-besta.it*

Guida breve all'editing secondo stili pre-impostati

Per utilizzare questo modello di documento è sufficiente posizionarsi nelle diverse sezioni e sostituire il testo di esempio con il proprio.

I numeri di pagina sono pre-ordinati e verranno aggiornati automaticamente.

Nelle pagine che seguono è stato impostato un esempio di documento nelle sue distinte parti: titolo, autori, paragrafi di testo, tavole, figure, riferimenti bibliografici. I testi utilizzati hanno esclusivamente uno scopo dimostrativo e possono essere modificati a seconda delle esigenze.

Qualora si ritenga utile comporre un indice delle parti del documento, può essere utilizzata la struttura appositamente predisposta. Essa si aggiorna posizionandosi sull'indice e tramite il tasto destro del mouse, attivando la voce “Aggiorna campo”: in tal modo si aggiorna la numerazione. Se si ritiene di intervenire sui titoli dei paragrafi, per aggiornare l'indice è sufficiente utilizzare l'opzione “Aggiorna intero sommario”.

In caso di stampa su carta, occorre selezionare dal menu di Word a sinistra, la funzione Stampa e, in seguito, il menu “Proprietà stampante”, la voce Layout e, infine, attivare la voce “Libretto” (stampa fronte/retro).

Pubblicazione non in vendita.
Nessuna riproduzione, traduzione o adattamento
può essere pubblicata senza citarne la fonte.
Copyright® PoliS-Lombardia

PoliS-Lombardia
Via Taramelli, 12/F - 20124 Milano
www.polis.lombardia.it

INDICE

INDICE.....	3
INTRODUZIONE	5
OBIETTIVI SPECIFICI E STRATEGICI DEL PROGETTO.....	8
METODOLOGIA ADOTTATA.....	9
DESTINATARI/BENEFICIARI DEL PROGETTO.....	12
DESCRIZIONE DEL PROGETTO, DELLE SUE FASI E TEMPISTICHE	13
COSTI DI IMPLEMENTAZIONE O REALIZZAZIONE.....	14
RISULTATI ATTESI	15
CONCLUSIONI	166
BIBLIOGRAFIA.....	17

INTRODUZIONE

L'**intelligenza artificiale (Artificial Intelligence, AI)** mira a creare processi che analizzano l'ambiente e a mettere in atto azioni per ottimizzare il successo verso un obiettivo predeterminato ([Sotoudeh et al., 2019](#)).

Machine learning è un sottotipo di AI focalizzata allo sviluppo di algoritmi che identificano pattern all'interno di dati senza informazioni preliminari. Questi algoritmi possono essere definiti come "supervised" (formati su un insieme di dati selezionati): in questo scenario è fondamentale la disponibilità di grandi quantità di dati. Nel caso di algoritmi "unsupervised" vengono identificati pattern nascosti in gruppi di dati non noti preliminarmente ai ricercatori. I dati possono essere numerici o nominali.

Gli algoritmi Random Forest e SVM (Support Vector Machines) sono due esempi di algoritmi supervised relativamente semplici utilizzabili per classificare un oggetto in differenti categorie. Artificial Neural Network (ANN) è invece un algoritmo più complesso di machine learning che in qualche modo tenta di mimare la funzionalità delle reti neurali.

Deep learning è una variante di ANN che aggiunge complessità usando strati multipli ("deep") di una rete neurale artificiale.

CNN (convolutional neural network) è un'altro tipo di ANN in cui i dati di imaging sono processati in rappresentazioni ridotte delle loro caratteristiche.

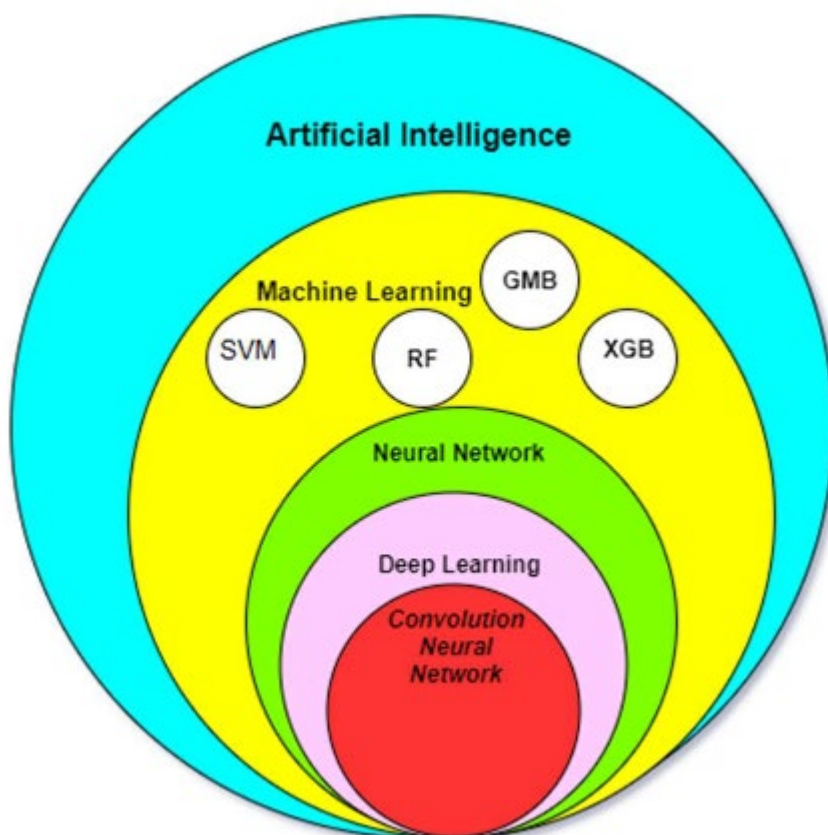


Figura 1. Relazione tra i più comuni metodi di AI in medicina. SMV, support vector machine; RF random forest algorithm; GMB, gradient boosting machine; XGB, XGboost (da Sotoudeh et al, 2019).

Nel marzo 2018 un lavoro su *Nature* dimostrava come l'uso di metodiche di machine learning potesse affinare significativamente la diagnostica dei tumori cerebrali basata su tecniche istologiche "tradizionali" (Capper *et al.*, 2018).

I tumori cerebrali sono un gruppo eterogeneo che include le metastasi cerebrali, che hanno l'incidenza nettamente maggiore, i meningiomi ed i gliomi (McFaline-Figueroa and Lee, 2018). I dati epidemiologici recenti USA valutano che l'incidenza dei tumori cerebrali primitivi (non metastatici) benigni e maligni sia di 23.4 nuovi casi/100.000 abitanti /anno (Ostrom *et al.*, 2019). I gliomi sono i più frequenti tumori primitivi maligni del cervello e tra essi il glioblastoma è il più aggressivo.

La classificazione dei tumori attualmente in vigore è del 2016. La tabella allegata dà un'idea della sua complessità.

Table 1 Primary Central Nervous System Tumors²⁵	
Diffuse astrocytic and oligodendroglial tumors	Embryonal tumors
Diffuse astrocytoma, IDH-mutant	Medulloblastomas, genetically defined
Diffuse astrocytoma, IDH-wild type	• WNT-activated
Diffuse astrocytoma, NOS	• SHH-activated and TP53 mutant
Anaplastic astrocytoma, IDH-mutant	• SHH-activated and TP53-wildtype
Anaplastic astrocytoma, IDH-wild type	• Non-WNT/Non-SHH
Anaplastic astrocytoma, NOS	Medulloblastomas, histologically defined
Glioblastoma, IDH-wild type	• Classic
Glioblastoma, IDH-mutant	• Desmoplastic/nodular
Glioblastoma, NOS	• With extensive nodularity
Diffuse midline glioma, H3 K27M-mutant	• Large cell/anaplastic
Oligodendroglioma, IDH-mutant and 1p/19q co-deleted	Medulloblastoma, NOS
Oligodendroglioma, NOS	Other embryonal tumors
Anaplastic oligodendroglioma, IDH-mutant and 1p/19q co-deleted	Tumors of the cranial and spinal nerves
Anaplastic oligodendroglioma, NOS	eg, Schwannoma, neurofibroma
Oligoastrocytoma*, NOS	Meningiomas
Anaplastic oligoastrocytoma*, NOS	Mesenchymal, nonmeningothelial tumors
Other astrocytic tumors	eg, Solitary fibrous tumor/hemangiopericytoma, hemangioblastoma
eg, Pilocytic astrocytoma	Melanocytic tumors
Ependymal tumors	Lymphomas
eg, Ependymoma	Diffuse large B-cell lymphoma of the central nervous system
Other gliomas	Immunodeficiency-associated central nervous system lymphomas
Choroid plexus tumors	• AIDS-related diffuse large B-cell lymphoma
Neuronal and mixed neuronal-gliar tumors	• EBV-positive diffuse large B-cell lymphoma, NOS
Tumors of the pineal region	Intravascular large B-cell lymphoma
eg, Pineocytoma, pineoblastoma	Other CNS lymphomas
Histiocytic tumors	Germ cell tumors
	Tumors of the sellar region
	Craniopharyngioma
	Granular cell tumor of the sellar region
	Pituicytoma
	Spindle cell oncocytoma

AIDS = acquired immunodeficiency syndrome; CNS = central nervous system; EBV = Epstein-Barr virus; H3 K27M = Histone H3 variant with lysine to methionine substitution at codon 27; IDH = isocitrate dehydrogenase; NOS = not otherwise specified, denotes a histological diagnosis in the absence of molecular information; SHH = Sonic hedgehog gene; TP = Tumor protein p53; WNT = Wnt signaling pathway; 1p/19q co-deleted = Co-deletion of the short arm of chromosome 1 and the long arm of chromosome 19.

*Oligoastrocytoma is a histological diagnosis. In the presence of molecular information, these tumors are re-classified as astrocytomas or oligodendrogliomas.

Tabella 1. Classificazione 2016 dei tumori cerebrali primitivi.

Tale complessità è una delle cause di possibili incertezze diagnostiche che possono avere una frequenza non trascurabile. Un lavoro del 1997 mostra che su 500 casi rivisitati da neuropatologi discrepanze diagnostiche serie o sostanziali fossero presenti in 140 casi (28%) ([Bruner et al., 1997](#)). In una analisi più recente 249 su 1281 casi rivisti presso l'NCCN (National Comprehensive Cancer Network) USA 249 (19.4%) non avevano avuto una diagnosi definitiva dalla istituzione di provenienza ([Bruner et al., 2017](#)). Nel caso dei tumori analizzati con l'intelligenza artificiale il computer non è riuscito ad assegnare una diagnosi in una frazione minore di casi (11.5%) e nel 12.6% aveva dato una diagnosi diversa da quella istologica, come riassunto nella figura sottostante, tratta da un commento al lavoro di Capper et al su *Nature*.

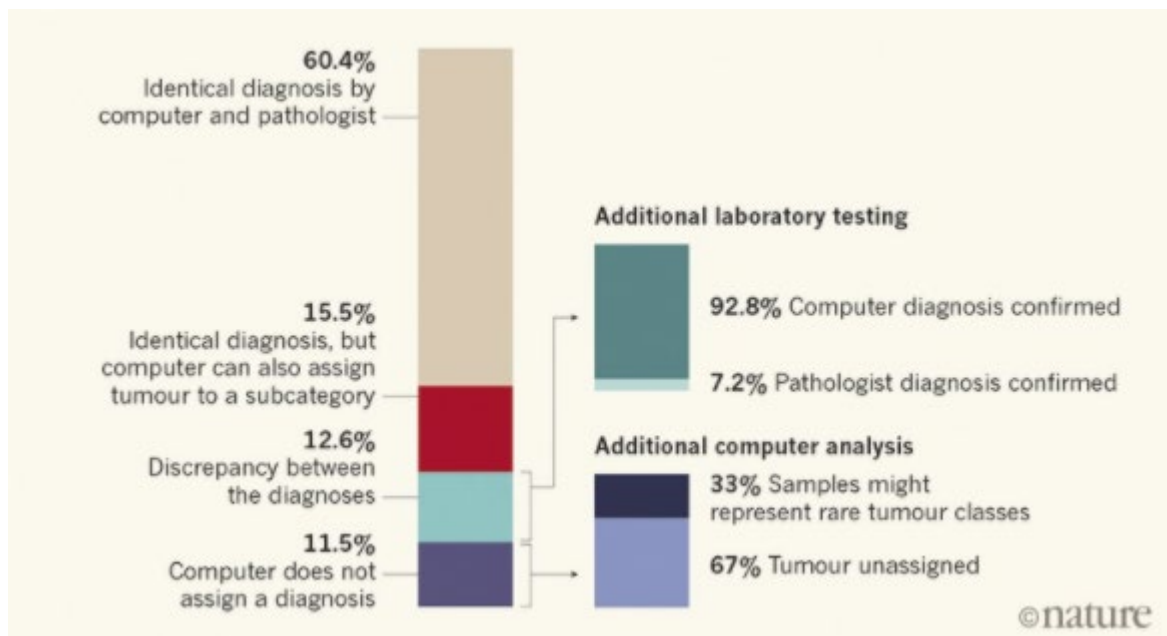


Figura 2. Classificazione dei tumori cerebrali usando un approccio di machine learning ([Wong and Yip, 2018](#)).

OBIETTIVI SPECIFICI E STRATEGICI DEL PROGETTO

L'obiettivo specifico del progetto è di sviluppare nell'ambito di un IRCCS lombardo una diagnostica dei tumori cerebrali basata su tecniche di AI. Nella fase pilota l'analisi può essere ristretta ai soli casi che due patologi esperti ritengano di diagnosi complessa e non univoca alla luce dei criteri WHO 2016.

L'obiettivo strategico è di estendere l'analisi ad una casistica più ampia creando uno o due centri HUB lombardi a cui centralizzare le diagnosi complesse in neuro-oncologia con l'ambizione di estendere tale servizio a livello nazionale.

METODOLOGIA ADOTTATA

La classificazione dei tumori cerebrali tramite machine learning si basa sui profili di metilazione del loro DNA. La metilazione del DNA implica l'aggiunta di un gruppo metile CH₃ al carbonio in posizione 5 della citosina (5-meC) ed è presente primariamente nel contesto del dinucleotide CpG in cellule eucariotiche (CpG island) (figura 3A).

I pattern di metilazione del DNA normale e tumorale possono variare significativamente (figura 3B) anche se i meccanismi che sottendono questi processi sono in larga parte ignoti.

..

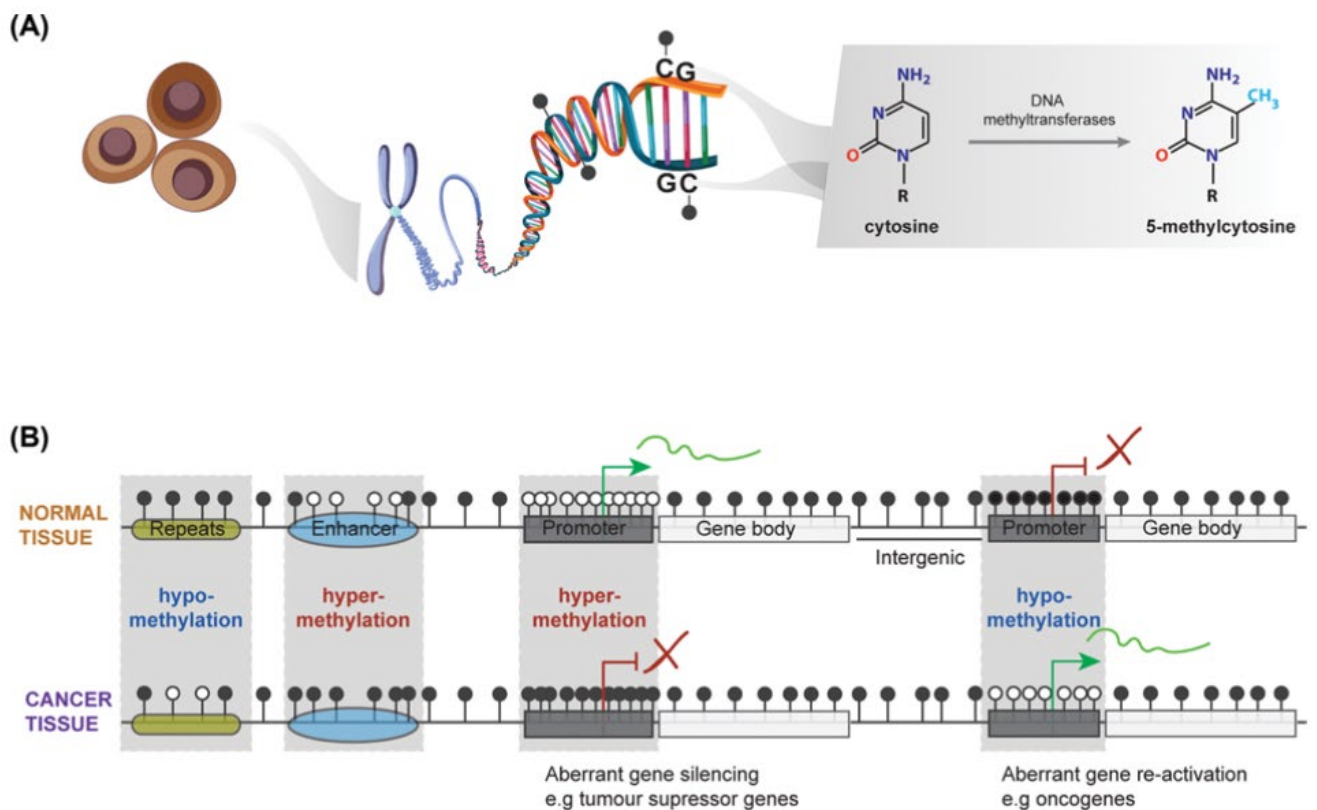


Figura 3. Il DNA di cellule normali o tumorali esibiscono diversi profili di metilazione del DNA ([Skvortsova et al., 2019](#)).

La metodologia implica l'estrazione del DNA da tessuto tumorale fissato in formalina e incluso in paraffina. L'utilizzo dell'estrattore Maxwell e del relativo kit di reagenti (Promega) sembrano garantire i migliori risultati in termini di qualità del DNA ([Jaunmuktane et al., 2019](#)).

Intelligenza artificiale per le diagnosi complesse in neuro-oncologia

Il passaggio successivo è la modificazione del DNA con bisulfite che a pH acido provoca la conversione della citosina in uracile (a sua volta convertito in timina dopo l'amplificazione per PCR), mentre la citosina metilata non si modifica.

Due methylation arrays sono disponibili presso Illumina, 450 k o 850 k (EPIC) per ottenere profili di metilazione dell'intero genoma. La strumentazione necessaria è NextSeq 550 o NextSeq 2000 di Illumina.

La sequenza sperimentale è illustrata nella figura 4.

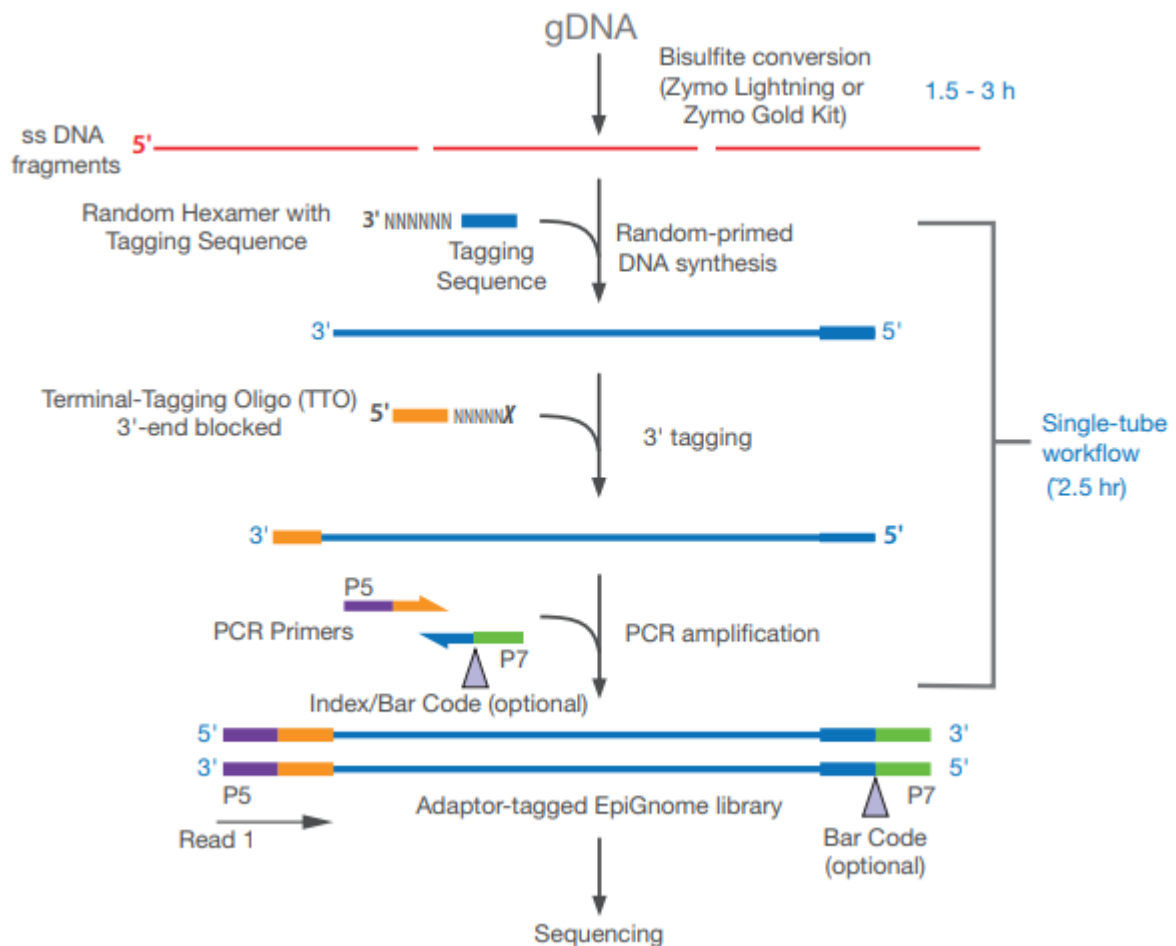


Figura 4. Workflow per la preparazione di libreria (documentazione Illumina).

L'analisi dei dati verrà poi sviluppata online sulla piattaforma disponibile presso:

<https://www.molecularneuropathology.org> (Università di Heidelberg, figura 5).

The screenshot shows the homepage of MolecularNeuropathology.org. At the top, there are logos for dkfz., UniversitätsKlinikum Heidelberg, KITZ Hopp-Kindertumorzentrum am NCT Heidelberg, and MolecularNeuropathology.org. Below the logos is a navigation menu with items: Home, Classifier, Reference set, Upload sample, Analysis, Contact, Legal notice, Data privacy, and a language selector (en). There are also 'Sign in' and 'Registration' buttons.

Welcome to MolecularNeuropathology.org - The platform for next generation neuropathology.

This website represents the access point for DNA methylation-based classification of central nervous system tumors. For the scientific background and interpretation of the data, please see Capper D, Jones DTW, Sill M, Hovestadt V et al., *Nature*. 2018 Mar 22;555(7697):469-474.

To implement the methylation profiling classifier you are required to generate and upload unprocessed IDAT-files of Illumina Human Methylation 450 BeadChip arrays or EPIC BeadChip arrays of your samples of interest. This data is then automatically compared to methylation data of a reference cohort comprising over 2800 neuropathological tumors of almost all known entities (currently over 80 tumor classes or subclasses included). Within a short time you will receive an E-Mail report of the methylation profiling of your case, a low resolution copy number plot calculated from your array data (useful e.g. for 1p/19q analysis or the detection of all sorts of amplifications and deletions) and an estimation of MGMT promoter methylation status.

Occasional updates may be required for either inclusion of new tumor classes or subtle changes of the EPIC array probe composition that may occur in a new batch. Older version will remain available.

Classification using methylation profiling is a research tool under development, it is not verified and has not been clinically validated. Implementation of the results in a clinical setting is in the sole responsibility of the treating physician.

Upload statistics:

IDAT files processed:	52567
Samples approved for classifier development:	38789
Statistics updated:	2020-10-14 12:43 UTC

Involved parties:

University Hospital Heidelberg
Neuropathology
Pediatric Oncology
Neurooncology
Neurosurgery
Radiation Oncology and Therapy

German Cancer Research Center (DKFZ)
Pediatric Neurooncology
CCU Neuropathology
Biostatistics
Molecular Genetics

Figura 5. Schermata iniziale di molecularneuropathology.org

DESTINATARI/BENEFICIARI DEL PROGETTO

Destinatari ideali del progetto sono gli IRCCS che ospitano un'area integrata di neuro-oncologia con tutte le specializzazioni necessarie: MRI per diagnosi, neurochirurgia, neuropatologia, radioterapia, neuro-oncologia medica, genetica molecolare (e.g. IRCCS Besta, IRCCS San Raffaele). Tale integrazione garantisce un flusso ottimale di attività per il paziente e quindi attrae un numero significativo di pazienti, tale da giustificare l'implementazione di un questo work-flow (al momento) sofisticato.

Beneficiari del progetto possono essere altri centri inizialmente lombardi e potenzialmente nazionali dove siano presenti strutture di neurochirurgia ed anatomia patologica con u volume adeguato di pazienti trattati/anno.

Beneficiari, in ultima analisi, saranno poi i pazienti per i quali la possibilità di una diagnosi più certa può comportare almeno informazioni prognostiche robuste e potenzialmente, l'accesso a terapie mirate, al momento infrequenti ma in verosimile crescita futura.

DESCRIZIONE DEL PROGETTO, DELLE SUE FASI E TEMPISTICHE

Possiamo ipotizzare una articolazione in cinque fasi per una durata complessiva di due anni.

1.1 Fase 1 (mesi 1-6)

Verifica della disponibilità di strumentazione e reagenti necessari. Messa a punto delle tecniche ottimali di estrazione del DNA e modificazione con bisolfito con risultati riproducibili usando campioni FFPE retrospettivi (n=15). Completamento di 5 run pilota su hardware NextSeq 550 (o NextSeq 2000).

1.2 Fase 2 (mesi 7-8)

Caricamento dati delle sequenze pilota sulla piattaforma www.molecularneuropathology.org. Definizione delle necessità in termini di collegamento informatico per il caricamento efficiente dei dati grezzi (raw data) generati e per la successiva analisi dei dati.

1.3 Fase 3 (mesi 9-12)

Verifica di congruità diagnostica istologia-dati da metilazione in casi noti (n=30).

1.4 Fase 4 (mesi 13-17)

Analisi di casi con diagnosi non univoca, come definito da due patologi (n≥15).

1.5 Fase 5 (mesi 18-24)

Estensione delle analisi a casi resi disponibili da centri esterni.

Le criticità principali possono essere incontrate nelle prime due fasi.

Per la fase 1, chiaramente il progetto presuppone un impegno finanziario con fondi da reperire tramite grant specifici (e.g. FRRB, AIRC, Fondazione CARIPLO, MdS-ricerca finalizzata).

Fase 2. Alla luce dell'esperienza pregressa è possibile si debba mettere in essere un collegamento di rete dedicato per questi trasferimenti dati.

COSTI DI IMPLEMENTAZIONE O REALIZZAZIONE

Il progetto presuppone la disponibilità di hardware adeguato (NextSeq550 o NextSeq2000, di Illumina, nello specifico).

Diversi dati convergono nell'indicare che l'estrazione di acidi nucleici di qualità sia un passaggio critico e che l'estrattore Maxwell RSC (Promega) garantisca tale qualità. Questo si può rilevare particolarmente rilevante volendo garantire in modo centralizzato la qualità di materiale provenienti da centri esterni. Il costo di Maxwell RSC è di 28.500 euro, IVA esclusa. Lo strumento può processare 16 campioni per volta con un turnaround time di 60 minuti.

Il costo chiave è quello dei reagenti. L'Infinium Methylation EPIC BeadChip kit permette di valutare 850,000 siti di metilazione a livello di singolo nucleotide. Un kit per l'analisi di 32 campioni ha un costo di 7725 euro, 241 euro a campione.

L'analisi bioinformatica è gratuita come pure lo storage dei dati. Da valutare tuttavia se sia necessaria l'arruolamento di un laureato in bioinformatica o se l'analisi semi-automatica dei dati non richieda questo livello di specializzazione.

Queste osservazioni sono in linea con il costo di 380 sterline (421 euro) per campione stimato nell'analisi del gruppo dell'University College di Londra (Jaunmuktane *et al.*, 2019) e di 550 dollari americani (468 euro) dell'Università di Toronto (Karimi *et al.*, 2019), strutture dove l'analisi del metiloma è stata implementata con successo.

A fronte di risultati positivi del progetto dopo i primi due anni (vedi "Descrizione del progetto") è ipotizzabile che la direzione strategica dell'IRCCS possa ad arrivare ad un accordo con la Direzione Generale Welfare per implementare stabilmente a livello regionale questo approccio diagnostico fortemente innovativo.

RISULTATI ATTESI

Il lavoro iniziale del gruppo di Heidelberg aveva portato ad un cambiamento di diagnosi nel 12% dei tumori analizzati (Capper *et al.*, 2018).

Nel report inglese la diagnosis è stata modificata in 45 su 179 casi, il 25% (Jaunmuktane *et al.*, 2019).

Nel report della University Health Network di Toronto i numeri complessivi sono meno comparabili, ma gli autori concludono che nel 15% dei 55 casi selezionati per l'analisi machine learning del metiloma a causa di diagnosi pregresse incerte vi è stato un impatto importante sulle decisioni cliniche, mirato ad evitare trattamenti non necessari o insufficienti (Karimi *et al.*, 2019).

Ci attendiamo quindi che, a fronte di costi contenuti, anche per la popolazione italiana potenzialmente in esame si possa arrivare a percentuali analoghe (15%) di pazienti cui possano essere offerti decisioni cliniche orientate razionalmente sulla base di dati.

CONCLUSIONI

Il progetto è partito dalla definizione generali tecniche di intelligenza artificiale applicabili in ambito bio-medico. L'approccio machine learning e l'utilizzo di un algoritmo "random forest" nell'ambito dell'analisi di DNA tumorale metilato può portare alla creazione di un classificatore in grado di migliorare la nostra capacità diagnostica. Lo schema è rappresentato nella figura 6.

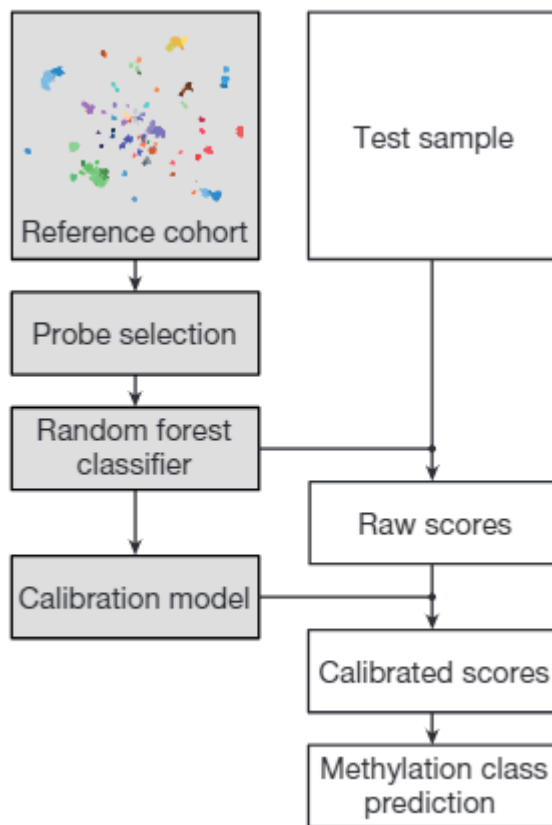


Figura 6. Schema dei principali componenti del classificatore (in grigio) e dei passaggi sperimentali partendo dal campione da analizzare (in bianco).
Da (Capper *et al.*, 2018).

Il progetto richiede la disponibilità di un budget iniziale contenuto. Le conferme ottenute in strutture indipendenti confermano la solidità di questo approccio e ne stimolano lo sviluppo anche nel nostro paese. E' infine da considerare che l'analisi AI di DNA tumorale metilato si potrebbe estendere a DNA cell-free tumorale nell'ambito della

espansione del ruolo della biopsia liquida in neuro-oncologia.

Dati recenti hanno dimostrato che lo studio del metiloma su DNA libero tumorale plasmatico di pazienti neuro-oncologici, con una nuova metodica che non prevede la conversione con bisolfito, può portare ad una discriminazione accurata del tumore intracranico d'origine (Nassiri *et al.*, 2020).

BIBLIOGRAFIA

- Bruner JM, Inouye L, Fuller GN, Langford LA. Diagnostic discrepancies and their clinical impact in a neuropathology referral practice. *Cancer* 1997; 79: 796–803.
- Bruner JM, Louis DN, McLendon R, Rosenblum MK, Archambault WT, Most S, et al. The utility of expert diagnosis in surgical neuropathology: analysis of consultations reviewed at 5 national comprehensive cancer network institutions. *J. Neuropathol. Exp. Neurol.* 2017; 76: 189–194.
- Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. *Nature* 2018; 555: 469–474.
- Jaunmuktane Z, Capper D, Jones DTW, Schrimpf D, Sill M, Dutt M, et al. Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre. *Acta Neuropathol. Commun.* 2019; 7: 24.
- Karimi S, Zuccato JA, Mamatjan Y, Mansouri S, Suppiah S, Nassiri F, et al. The central nervous system tumor methylation classifier changes neuro-oncology practice for challenging brain tumor diagnoses and directly impacts patient care. *Clin. Epigenetics* 2019; 11: 185.
- McFaline-Figueroa JR, Lee EQ. Brain Tumors. *Am. J. Med.* 2018; 131: 874–882.
- Nassiri F, Chakravarthy A, Feng S, Shen SY, Nejad R, Zuccato JA, et al. Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. *Nat. Med.* 2020; 26: 1044–1047.
- Ostrom QT, Cioffi G, Gittleman H, Patil N, Waite K, Kruchko C, et al. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro Oncol.* 2019; 21: v1–v100.
- Skvortsova K, Stirzaker C, Taberlay P. The DNA methylation landscape in cancer. *Essays Biochem.* 2019; 63: 797–811.
- Sotoudeh H, Shafaat O, Bernstock JD, Brooks MD, Elsayed GA, Chen JA, et al. Artificial intelligence in the management of glioma: era of personalized medicine. *Front. Oncol.* 2019; 9: 768.
- Wong D, Yip S. Machine learning classifies cancer. *Nature* 2018; 555: 446–447.

