



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia Generale

**Corso di laurea in Neuroscienze e Riabilitazione
Neuropsicologica**

Elaborato finale

*Modelli di Machine Learning per la predizione di attacchi
epilettici sulla base del segnale EEG*

*Machine Learning models for epileptic seizures prediction based on EEG
signal*

Relatore: Prof. Alberto Testolin

Laureando: Nicholas Moro

Anno Accademico 2021/2022

INDICE

1. INTRODUZIONE	5
1.1 Epilessia	5
1.1.1 Epilessia Generalizzata	7
1.1.2 Epilessia Focale	9
1.1.3 Cenni di Terapia	10
1.2 Elettroencefalogramma	12
1.2.1 Preprocessing del segnale elettroencefalografico	15
1.2.2 Detezione e diagnosi di crisi epilettiche	19
1.2.3 Predizione di crisi epilettiche	21
2. DEEP LEARNING	23
2.1 Dataset e iperparametri	27
2.2 Reti neurali convoluzionali	29
2.3 Transfer Learning	35
3. METODOLOGIA	37
3.1 VGG 16	39
3.2 Vanilla CNN	43
4. RISULTATI	45
4.1 VGG 16	45
4.2 Vanilla CNN	49
5. DISCUSSIONE	54
BIBLIOGRAFIA	62

ABSTRACT

L'epilessia è una malattia neurologica cronica caratterizzata da convulsioni ripetute e non provocate e con un tasso di prevalenza relativamente alto. Le convulsioni possono essere gestite con l'aiuto di farmaci antiepilettici e altre opzioni di trattamento come l'intervento chirurgico. Sfortunatamente, le persone con epilessia farmaco-resistente non sono in grado di ottenere la piena libertà poiché non esiste un trattamento efficace per gestire il tipo di crisi o la sua causa. La possibilità di previsione dell'attacco prima del suo inizio può rivelarsi cruciale soprattutto in questi pazienti. Su questo presupposto si basa la ricerca, argomento di trattazione di questa tesi, svolta in team multidisciplinare composto da uno studente di Ingegneria per le comunicazioni multimediali e internet, uno studente di Physics of Data, uno studente di Neuroscienze e Riabilitazione Neuropsicologica ed un dottorando in Brain, Mind and Computer Science.

Attraverso il rilevamento di convulsioni da lunghe registrazioni EEG l'obiettivo del progetto è quello di esplorare due differenti modelli di deep learning con lo scopo di predire un attacco epilettico. Dopo aver opportunamente pre-processato il segnale, gli spettrogrammi del tracciato antecedente la crisi epilettica sono stati suddivisi in quattro categorie. Il dataset composto dalle immagini del segnale EEG, è stato utilizzato per addestrare e testare in un compito di classificazione una "VGG 16" mediante un'operazione di transfer learning e una "Vanilla CNN" (implementata dallo studente del corso di laurea in Physics of Data) con lo scopo di categorizzare gli spettrogrammi entro le quattro classi.

1. INTRODUZIONE

1.1 Epilessia

L'epilessia è la malattia che causa crisi comiziali (seizures), le quali sono un'alterazione dello stato di coscienza accompagnata o meno da segni/sintomi neurologici o psichiatrici. Il termine crisi comiziale include tutte le alterazioni parossistiche della normale attività elettrica del cervello.

Ecco, quindi, che una crisi epilettica è definibile come una scarica neuronale elettrica anomala ed incontrollata che interrompe transitoriamente il normale funzionamento cerebrale.

Una crisi in genere provoca alterazioni dello stato di coscienza, movimenti anormali, convulsioni o stati mentali alterati. Con il termine "convulsione" si fa riferimento a crisi comiziali motorie (Shoka et al., 2019).

È bene evidenziare che normalmente l'attivazione e quindi l'eccitabilità dei neuroni corticali è modulata da circuiti di feedback eccitatori e inibitori tra loro in equilibrio e presenti sia in aree adiacenti (connessioni laterali) sia a livello profondo (si tratta di circuiti che dalla corteccia vanno ai gangli della base e da qui al talamo). In questo circuito ipotetico, uno stimolo attiva una determinata popolazione neuronale corticale che dà una risposta, la quale, però, nel tempo si spegne proprio grazie ai feedback negativi presenti (Angeles et al., 1981).

Invece nel caso di un pattern di scarica di tipo epilettico, una stimolazione ripetitiva della corteccia cerebrale (e.g. guardare la televisione o giocare a un videogioco, casi in cui si ha una stimolazione a 10-40 Hz dello schermo), causa l'attivazione di questa peculiare popolazione neuronale che però, a causa di difetti cellulari o strutturali, non viene più controllata dai circuiti di feedback inibitori.

Ecco che l'attivazione corticale ripetitiva potenzia i circuiti eccitatori e sopprime quelli inibitori, creando un circolo che si autoalimenta facilitando l'eccitazione dei neuroni corticali adiacenti (crisi comiziale focale) e può interessare anche il talamo e i suoi sistemi di proiezione diffusi alla corteccia cerebrale dando luogo, in questo caso, a una crisi comiziale generalizzata (Ko & Benbadis, 2016).

La crisi può anche partire da circuiti sottocorticali che attivano direttamente e in maniera diffusa la corteccia cerebrale.

L'epilessia è la patologia neurologica cronica che colpisce 50 milioni di persone in tutto il mondo. L'incidenza varia da 40 a 70 ogni 100.000 abitanti/anno nei paesi industrializzati e da 100 a 190 ogni 100.000 abitanti/anno nei paesi in via di sviluppo (Smith, 2005). Secondo uno studio epidemiologico del 1996, la curva di incidenza specifica per età ha una distribuzione bimodale con un picco nel primo anno di vita, seguito da una riduzione nell'età media ed una successiva crescita nella tarda età (80 anni) (Savarino et al.,2012).

In letteratura sono presenti numerose classificazioni delle diverse tipologie di crisi epilettiche. Come riportato nella tabella 1.1, in questo lavoro si prenderà spunto dall'International League Against Epilepsy (ILAE) per presentare i principali cluster epilettici (Cherian & Kanaga, 2022).

Essenzialmente le crisi si dividono in parziali (dette anche focali) e crisi generalizzate sulla base del meccanismo neurologico attivante.

Nelle forme focali l'idea è che vi sia un focolaio corticale, mentre a livello generalizzato si pensa a un focolaio tipicamente a livello talamico (ci sono in realtà anche delle forme generalizzate con partenza da un focolaio corticale che si generalizza molto rapidamente). Nel primo caso la crisi resta limitata a una parte della corteccia e con essa la manifestazione clinica dei sintomi, che dipende dall'area interessata; nel secondo e terzo caso invece la crisi è bilaterale e di conseguenza ci saranno sintomi bilaterali o di perdita di coscienza.

EPILESSIA GENERALIZZATA	EPILESSIA FOCALE
Grande male	Semplice
Piccolo male	
Sindrome di West	Parziale complessa
Epilessia mioclonica giovanile	

Tabella 1.1 Classificazione epilessia

1.1.1 Epilessia Generalizzata

Le crisi generalizzate si suddividono in crisi tonico-cloniche (grande male), assenza (piccolo male), sindrome di West, epilessia mioclonica giovanile.

Le crisi generalizzate tonico-cloniche si distinguono in una fase preclinica e tre fasi cliniche: tonica, clonica, periodo post-ictale.

Nella fase preclinica il paziente può manifestare (per ore o minuti) un malessere generalizzato, astenia, difficoltà nel pensiero, apatia, depressione, irritazione, accelerazione della parola o del pensiero, sintomi psichiatrici etc. La presenza di questi prodromi, soprattutto nei pazienti che vanno incontro a crisi di grande male con una certa periodicità, viene correttamente interpretata dallo stesso paziente come un evento che anticipa la crisi vera e propria.

Quando questa fase termina e lascia spazio alla fase tonica solitamente si può osservare un cambiamento del paziente come un urlo o una perdita di coscienza con caduta a terra.

La fase tonica ha una durata di circa 10-20 secondi ed è caratterizzata da un'attività contrattile tonica che può manifestarsi con uno spasmo cronico tonico, con il morso della lingua, con la presenza di apnea a causa dello spasmo dei muscoli respiratori, pupille midriatiche e non responsive allo stimolo luminoso, svuotamento della vescica. In questa breve fase l'EEG è rapido e ad alta frequenza.

La fase clonica ha una durata compresa tra 30 secondi e 60 minuti ed è fondamentalemente un'alternanza di eccitazione e inibizione: lentamente la fase tonica evolve in una fase clonica caratterizzata da contrazioni simmetriche ai 4 arti. Nella fase clonica inizialmente si identificano dei tremori generalizzati (8-4 Hz) con successiva flessione violenta, un aumento dell'attività autonoma (aumento di frequenza cardiaca, BP, pressione) e un mantenimento dell'apnea fino alla fine della fase clonica, momento in cui spesso si verifica un profondo inspiro.

L'EEG è in fase di recupero ed è caratterizzato da onde ad alta frequenza alternate a delle pause che riflettono l'entrata in azione di meccanismi inibitori che cercano di fermare la crisi.

In ultimo, nel periodo post-ictale il paziente è non responsivo e può verificarsi incontinenza urinaria e perdita di saliva. L'EEG in questa fase è caratterizzato dalla risoluzione della crisi e da onde a basso voltaggio caratteristiche di un ipometabolismo.

L'assenza (piccolo male) è un'altra forma di epilessia generalizzata che tipicamente viene rilevata soprattutto nei bambini. Generalmente si manifesta con un'interruzione dell'attività che il soggetto sta compiendo, assenza di risposta, fissazione dello sguardo, occhi sbarrati, incoscienza, amnesia rispetto all'evento. Le crisi durano solitamente fino a 10 s. L'EEG si caratterizza da un'attività a 3 cicli al secondo (3 Hz), dunque da onde delta (0 - 4 Hz) molto alte e frequenti; le alterazioni EEG sono molto più precoci rispetto all'insorgenza dei sintomi.

La sindrome di West consiste in una triade caratteristica: spasmi infantili, arresto dello sviluppo psicomotorio e ipsaritmia; anche se uno dei tre elementi può essere omesso. Gli spasmi possono essere in flessione o in estensione, interessano collo, tronco ed arti, con scosse rapide e/o con caduta del capo; si presentano ripetuti in serie e più spesso all'addormentamento o al risveglio. L'età di esordio interessa il periodo tra il quarto ed il settimo mese e, comunque, prima dell'anno; l'incidenza viene stimata intorno a 1 su 5000. L'EEG è caratterizzato da una attività disorganizzata (ipsaritmia) costituita da onde lente, punte ed onde angolari di grande ampiezza, distribuite in modo anarchico.

L'epilessia giovanile mioclonica (o mioclono giovanile) è la forma di epilessia generalizzata idiopatica più comune nei giovani adulti e nei bambini più grandi; insorge tipicamente nell'adolescenza attorno ai 15 anni.

Questa forma di epilessia è caratterizzata dal mioclono, movimento che può essere unilaterale o bilaterale, ma spesso può presentare anche crisi generalizzate tonico-cloniche al risveglio a causa di miocloni che insorgono al mattino e coinvolgono tutto il corpo.

Il mioclono giovanile è una forma di epilessia benigna in quanto non dà compromissione progressiva intellettiva e tende a migliorare con il tempo.

Il pattern EEG tipico presenta onde theta (4 - 6 Hz) e un'attività poli-spike che si verifica in concomitanza del mioclono o che può essere intercorrente nei momenti in cui non si identifica un'attività comportamentalmente epilettica (Battistin, 2022).

1.1.2 Epilessia Focale

L'epilessia focale si suddivide in focale semplice (con coscienza preservata) e focale complessa (con alterazione dello stato di coscienza).

Le epilessie focali semplici possono essere di vari tipi in quanto qualsiasi parte della corteccia può essere focus epilettico. Ecco, quindi, che l'epilessia focale ha una sintomatologia che riflette la funzione dell'area interessata e del network associato ad essa. Dal punto di vista semeiologico si possono avere: sintomi somato-sensoriali (parestesie), sintomi focali motori (tic motori, movimenti tonico-clonici), ammiccamenti e movimenti anormali del viso, fenomeni autonomici (sudorazione, flushing, pallore), allucinazioni visive e uditive.

Le crisi complesse o con alterazione dello stato di coscienza tipicamente partono da zone associative del cervello dove ci possono essere tipicamente delle sindromi comportamentali più complicate. I sintomi associati sono quindi: allucinazioni visive o auditive formate, fenomeni psicologici o psichiatrici (reazioni di paura, deliri religiosi, attacchi di panico), alterazioni motorie complesse nel caso di iperattività frontale.

Infine vi è una terza categoria di sindromi epilettiche definite speciali, tra cui troviamo; deficit psichiatrici in pazienti epilettici, SUPED, epilessia Rolandica, crisi febbrili, crisi psicogeniche o pseudo-seizures (Battistin, 2022).

1.1.3 Cenni di Terapia

Il trattamento dell'epilessia comprende essenzialmente tre vie: farmaci, chirurgia, stimolazione del nervo vago (Battistin, 2022).

I farmaci principali sono farmaci anticonvulsivanti tra cui, ad esempio, la Carbamazepina e Levetiracetam che agiscono a livello pre-sinaptico. In particolare il primo ha un meccanismo d'azione che inibisce i canali Na^+ voltaggio-dipendenti mentre il secondo agisce sul legame delle vescicole VV2 presinaptiche (figura 1.1).

D'altro canto, il trattamento chirurgico è indicato se il paziente ha un'epilessia focale con una zona epilettogena che sia ben determinata e lontana da altre aree funzionali.

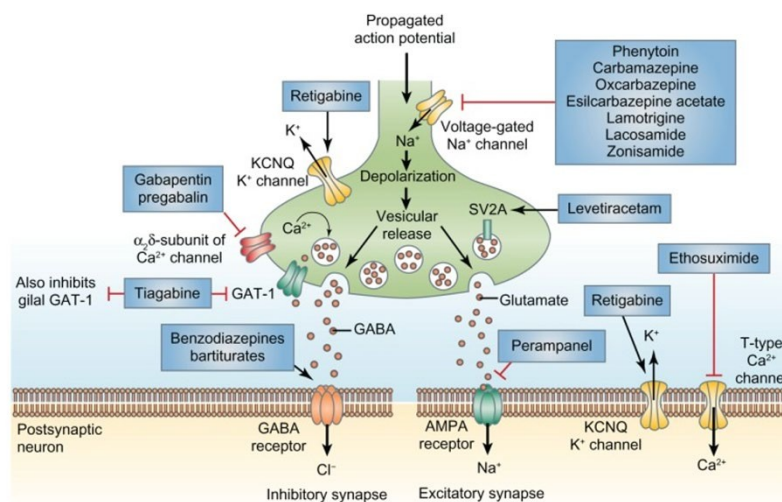


Figura 1.1 Meccanismo d'azione farmaci anti epilettici

Qualora i pazienti siano ineligibili alla chirurgia o la rifiutino, la stimolazione del nervo vago è una tecnica palliativa che non guarisce dalle crisi ma incrementa la qualità di vita. La stimolazione vagale è una procedura poco invasiva che viene attuata al nervo vago di sinistra lungo il suo decorso cervicale. Attorno a questa porzione del vago vengono applicati tre elettrodi che si avvolgono al tronco del nervo, collegati ad un generatore di stimoli posizionato nella regione sottocutanea clavare sinistra. La stimolazione antidromica (verso l'encefalo) induce l'attivazione di strutture quali il nucleo del tratto solitario e del rafe, attive nella secrezione di serotonina; quest'ultima agisce alzando la

soglia epiletto-gena di tutta la corteccia cerebrale, riducendo così la predisposizione ad avere attacchi epilet-tici. Questa metodologia dev'essere comunque accompagnata dal trattamento farmacologico.

1.2 Elettroencefalogramma

L'elettroencefalogramma (EEG) fu ideato da Hans Berger nel 1929 ed è una metodologia di imaging funzionale non invasiva usata per misurare l'attività elettrica della corteccia cerebrale a partire dal cuoio capelluto (Rasheed et al., 2020).

La corteccia cerebrale ha uno spessore di 2-4mm e vi si differenziano 6 diversi strati. Ogni strato si distingue sia per la tipologia di neuroni da cui è popolato, sia per il modo in cui è connesso con le aree del cervello. Sono le cellule piramidali del III e V strato ad essere maggiormente coinvolte nella genesi del segnale di EEG in quanto possiedono un corpo triangolare e un dendrite apicale che spesso raggiunge gli strati superficiali della corteccia.

L'attività post-sinaptica di queste cellule crea quindi il segnale utile registrabile sullo scalpo dall'elettrodo. Ogni dendrite è infatti una sorgente di campo elettrico ed essendo i diversi dendriti disposti parallelamente si può misurare un campo elettrico complessivo dato dalla somma dei diversi contributi.

Il segnale EEG è generato solamente dall'attività post-sinaptica e non dal potenziale d'azione che si propaga lungo l'assone.

Infatti, mentre i potenziali d'azione sono asincroni per la maggior parte degli assoni che corrono in molte direzioni rispetto alla superficie e quindi danno un contributo trascurabile all'acquisizione, i potenziali post-sinaptici (eccitatori e inibitori) sono sincronizzati per gruppi di cellule e la loro localizzazione non casuale permette di registrare un'attività utile.

La registrazione dei potenziali elettrici dei dendriti dei neuroni vicino alla superficie corticale è affidabile in tempo reale in quanto può effettuare misurazioni ogni millesimo di secondo. Infatti, la risoluzione temporale dei segnali EEG è decisamente migliore della risoluzione spaziale. Quest'ultima è bassa a causa dello scarso rapporto segnale-rumore (dipendente sia da fattori oggettivi quali rumori ambientali, ostruzione del cranio, ecc. che soggettivi come lo stato di affaticamento del soggetto) e degli effetti di sfocatura (il segnale EEG deve attraversare diversi strati dell'encefalo e della scatola cranica, entrambi buoni isolanti elettrici).

A seconda che l'attività del cervello umano venga acquisita utilizzando il cuoio capelluto o gli elettrodi intracranici, si distinguono due tipi di EEG: l'EEG di superficie o iEEG. L'ampiezza del segnale varia da 10 μV a 100 μV mentre la frequenza oscilla tra 1 Hz e 100 Hz (Bear et al., 2016).

Gli elettrodi vengono posizionati sul cuoio capelluto utilizzando il Sistema Internazionale 10-20 che individua alcune distanze costanti utilizzando tre specifici riferimenti anatomici: nasion, inion e punti pre-auricolari. Sulla base di questi tre punti si crea un sistema di riferimento tracciando fra questi delle linee fondamentali quali la sagittale antero-posteriore, mediale e laterale, la linea coronale frontale, centrale e parietale.

La rete che viene a formarsi attraverso l'intersezione di queste linee permette di individuare l'esatta posizione di ciascun elettrodo. La distanza fra un elettrodo e un altro è pari al 10 o 20% della lunghezza complessiva di ogni linea (da qui il nome di sistema internazionale 10-20) e questo assicura una buona copertura totale (Figura 1.2).

Questo sistema permette la disposizione dei vari elettrodi indipendentemente dalla forma individuale della testa.

Ogni posizione è indicata da una combinazione di una lettera e un numero: le lettere F, T, C, P e O indicano rispettivamente i lobi frontale, temporale, centrale, parietale e occipitale; i numeri pari 2, 4, 6 e 8 si riferiscono all'emisfero destro e i numeri dispari 1, 3, 5, 7 all'emisfero sinistro della corteccia cerebrale. Z rappresenta l'elettrodo posto al centro.

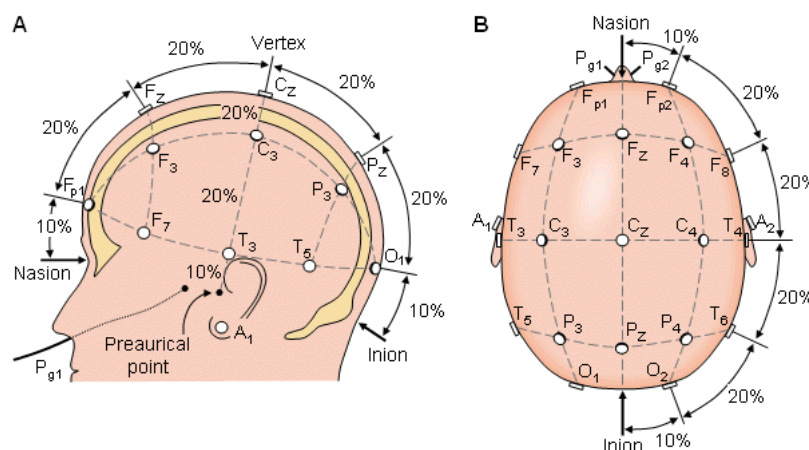


Figura 2.2 Sistema di posizionamento 10-20

Come mostrato in figura 1.2, il posizionamento può coinvolgere 21 elettrodi (sistema 10-20), 64 elettrodi (sistema 10-10), 128 elettrodi (sistema 10-5).

È bene evidenziare che oltre agli elettrodi per la misurazione dell'EEG vengono spesso aggiunti elettrodi addizionali per la misura di ECG (elettrocardiogramma), EOG (elettrooculogramma) e EMG (elettromiografia) della palpebra e dei muscoli che circondano gli occhi (utili soprattutto in compiti di visual search).

Questi rappresentano un segnale importante da acquisire perché, essendo molto più forti del segnale elettrico generato dalle cellule potenziali, costituiscono un fattore di disturbo; è quindi importante studiarne le caratteristiche per sottrarle alle informazioni dell'EEG di interesse.

1.2.1 Preprocessing del segnale elettroencefalografico

Come sottolineato precedentemente, nell'EEG l'attività elettrica originata nel cervello viene misurata al cuoio capelluto, quindi la qualità del segnale non solo diminuisce ma diverse informazioni spaziali possono andare perse. Al fine di porre rimedio a tale situazione, i segnali cerebrali vengono raccolti e i dati EEG grezzi vengono pre-elaborati per rimuovere il rumore e gli artefatti indesiderati.

Tale tecnica, chiamata preprocessing, è una procedura che permette di trasformare il segnale grezzo in un formato adatto per le successive analisi. Nel caso dei dati EEG la rimozione del rumore permette di incrementare il SNR (signal to noise ratio) e, quindi, di avvicinarsi al “vero” segnale EEG.

Gli artefatti sono segnali indesiderati presenti nell'EEG che possono portare a modifiche del segnale di interesse, incrementando il rumore presente nel dato; possono suddividersi in artefatti esterni o interni.

I primi provengono da fonti ambientali o esterne e sono generalmente causati da cavi di alimentazione, guasti agli elettrodi, rumore e movimenti della linea elettrica.

Gli artefatti interni invece sono relativi alle componenti fisiologiche e si distinguono principalmente in oculari, muscolari e cardiaci (Jiang et al., 2019).

L'origine degli artefatti oculari è dato dalla saccade (movimento orizzontale) e blinking (movimento verticale) che possono diffondersi al cuoio capelluto ed essere registrati dall'attività EEG.

Inoltre, a causa dell'effetto di conduzione del volume, sia l'artefatto oculare che l'attività EEG sono propagati sulla superficie della testa e registrati dagli elettrodi. Tali segnali oculari possono essere registrati utilizzando l'elettrooculogramma (EOG).

Gli artefatti muscolari possono essere causati da qualsiasi vicinanza muscolare ai siti di registrazione del segnale: contrazione, fonazione, deglutizione, ecc.

Gli artefatti muscolari misurati mediante EMG hanno una distribuzione ampia (da 0 Hz a > 200 Hz) che è influenzata dal grado di contrazione e allungamento muscolare; per tale motivo sono particolarmente difficili da eliminare.

Gli artefatti cardiaci possono essere introdotti quando gli elettrodi sono posizionati sopra o vicino ad un vaso sanguigno, dove il movimento di espansione e contrazione è dovuto al cuore (artefatti del polso); la loro frequenza è di circa 1,2 Hz.

Un'altra attività cardiaca nota come ECG misura il segnale elettrico emesso dal cuore. Contrariamente agli artefatti del polso, l'ECG può essere misurato con uno schema regolare caratteristico e registrato separatamente dall'attività cerebrale, quindi la sua rimozione può essere più semplice utilizzando una forma d'onda di riferimento. Le principali forme d'onda degli artefatti interni sono mostrati in figura 1.3.

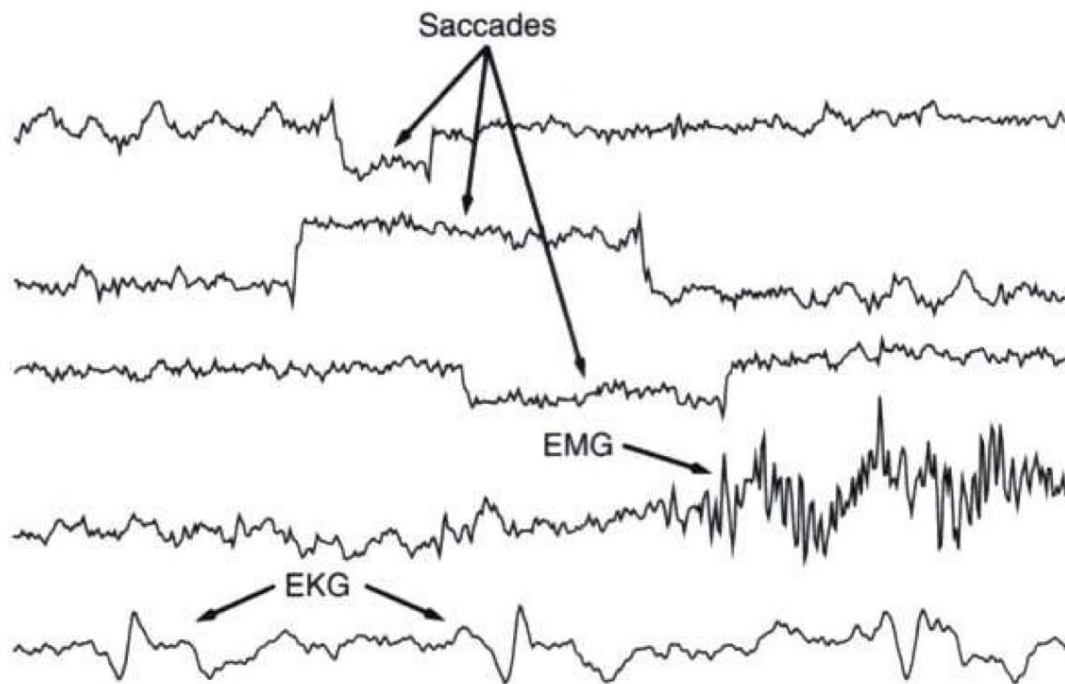


Figura 1.3 Forme d'onda tipiche degli artefatti interni

Poiché la pre-elaborazione rimuove le componenti indesiderate, la qualità del segnale EEG ne risulta migliorata. I filtri passa basso, passa alto, passa banda e notch sono la soluzione principale per rimuovere gli artefatti.

Il filtro passa basso è usato per rimuovere i rumori ad alta frequenza, elimina tutte le frequenze al di sopra di un determinato limite mantenendo quelle al di sotto del valore prefissato; il filtro passa alto, invece, rifiuta tutte le frequenze al di sotto di un determinato limite mantenendone quelle al di sopra. Questi primi due filtri sono spesso utili per rimuovere gli artefatti da movimento.

I filtri passa banda rappresentano una combinazione tra i primi due: trasmettono tutte le frequenze tra un limite di frequenza inferiore e superiore definiti.

Infine, il filtro Notch rimuove una singola frequenza ed è utile per ridurre il rumore di rete.

Un'altra tecnica cardine nel preprocessing è il downsampling.

Ad esempio, se in un sistema EEG con 64 canali e una frequenza di campionamento di 600 campioni al secondo (600 Hz) rappresentassimo ogni campione come un float a 32 bit, questo equivale a $(64 * 600 * 32) = 1.228.800$ bit al secondo o 150 kb/sec di dati.

È qui che entra in gioco il downsampling, una tecnica per ridurre il numero di campioni utilizzati, pur preservando le informazioni necessarie. Campionando un segnale fisiologico che presenta alte frequenze con una frequenza di campionamento troppo bassa si creano frequenze fantasma che nulla hanno a che vedere con il segnale di interesse (fig. 1.4). Per evitare questo fenomeno detto aliasing è necessario utilizzare filtri passa-basso per poter campionare a frequenze non troppo alte (riducendo la dimensione del file).

I filtri passa-basso anti-aliasing incorporati nei sistemi integrati evitano distorsioni e la comparsa di frequenze fantasma tagliando tutti i segnali ad alta frequenza endogeni e ambientali (es. EMG, neon, alimentatori switching, schermi CRT-LCD, ecc).

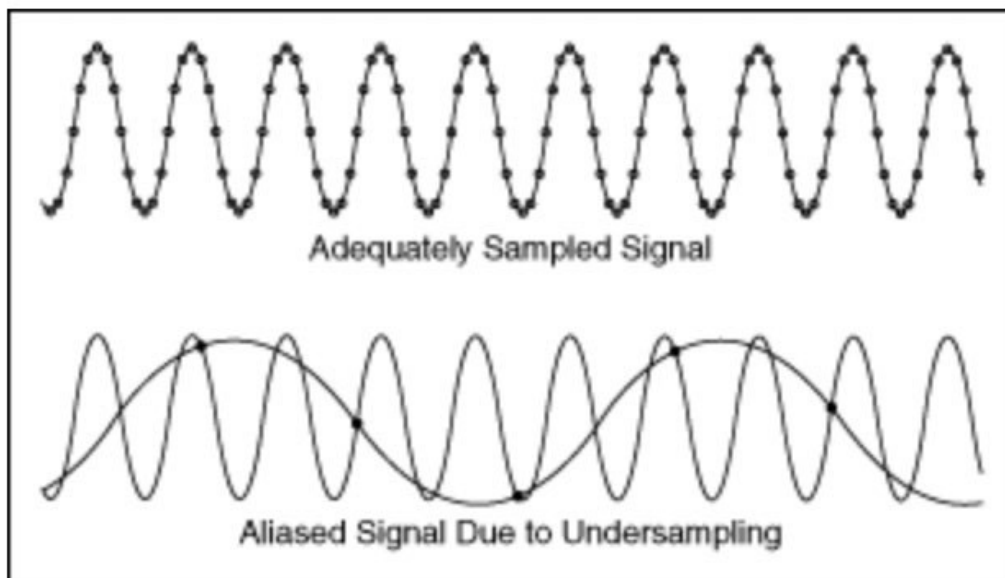


Figura 1.4 Frequenze fantasma dovute al fenomeno di aliasing

Per scegliere la corretta frequenza di campionamento ed il corrispettivo filtro anti-aliasing è utile tenere in considerazione il Criterio di Nyquist secondo cui la frequenza di campionamento deve essere più di 2 volte maggiore della frequenza massima presente nel segnale. Per non distorcere il segnale e perderne la forma, la frequenza di campionamento dovrebbe essere compresa tra 2 e 5 f_{max} . Il filtro dovrebbe quindi tagliare tutte le frequenze al di sotto di almeno la metà della frequenza di campionamento.

1.2.2 Detezione e diagnosi di crisi epilettiche

Per quanto concerne le tecniche di screening nella diagnosi delle crisi epilettiche vi sono due importanti categorie: neuroimaging funzionale e neuroimaging strutturale. La modalità di neuroimaging funzionale fornisce importanti informazioni sulla funzione cerebrale durante l'insorgenza di crisi epilettiche mentre la modalità strutturale fornisce ai medici informazioni sostanziali sulla struttura cerebrale degli individui con crisi epilettiche.

Le più importanti tecniche di neuroimaging funzionale sono l'EEG, la magnetoencefalografia (MEG), la tomografia a emissione di positroni (PET), la tomografia computerizzata a emissione di fotoni singoli (SPECT), la risonanza magnetica funzionale (fMRI), elettrocorticografia (ECoG) e spettroscopia funzionale nel vicino infrarosso (fNIRS) (Bear et al., 2016).

L'EEG è la tecnica “gold standard” per la diagnosi e previsione dell'epilessia in quanto possiede diverse caratteristiche che lo rendono preferibile da utilizzare per la ricerca. Insieme alla sua capacità di tracciare i vari cambiamenti che si verificano nel cervello durante l'epilessia, possiede un costo relativamente basso che lo rende in grado di essere utilizzato per un gran numero di pazienti e di registrare per una durata maggiore (Smith, 2005).

L'elettroencefalografia è utile nella diagnosi delle crisi epilettiche, in particolare dello stato epilettico parziale complesso e di quello dell'assenza. L'EEG può individuare anomalie legate all'epilessia (picchi, onde acuminate, complessi a punta e punta-onda, onde lente). Negli individui con crisi epilettiche primarie generalizzate, le anomalie epilettiche possono essere bilaterali, simmetriche e sincroniche, ma nei pazienti con crisi parziali possono essere localizzate.

Possibili risultati dell'elettroencefalografia possono comprendere:

- anomalie epilettiformi interictali (tra due crisi) nelle foci del lobo temporale in crisi parziali complesse con origine nel lobo temporale
- scoppi (burst) simmetrici bilaterali interictali di attività epilettiforme a 4-7 Hz in crisi tonico-cloniche primariamente generalizzate
- scariche epilettiformi focali in crisi secondariamente generalizzate

- scariche di punte e onde lente bilaterali a una frequenza di 3/secondo e generalmente normale attività elettroencefalografica di fondo in crisi di assenza tipiche
- fast ripples aperiodici ad elevata frequenza (250 – 600 Hz)
- scariche di punte e onde lente, solitamente a una frequenza < 2,5/secondo, associate in genere a una disorganizzazione interictale dell'attività di fondo e onde lente diffuse, in crisi di assenza atipiche
- poli-punte bilaterali e anomalie delle onde a una frequenza da 4 a 6 Hz nell'epilessia mioclonica giovanile

Tuttavia, la diagnosi deve comunque essere effettuata clinicamente poiché, ad esempio, se le convulsioni sono frequenti, è meno probabile che l'elettroencefalografia rilevi anomalie. Solo il 30-55% dei pazienti con epilessia nota può avere un'anomalia identificata da un primo EEG, e fino all'80-90% di questi pazienti può presentare anomalie identificate da successivi elettroencefalogrammi.

Talvolta i pazienti epilettici vengono monitorati 24 ore al giorno per una settimana in clinica con EEG a diversa frequenza (16 o 32 elettrodi) e con una video-sorveglianza (si parla di video-EEG). Questo monitoraggio viene fatto in presenza di infermieri specializzati nel rilevare e nel trattare una crisi epilettica.

I video-EEG possono servire nei casi di preparazione alla chirurgia per mappare in maniera più precisa la sede della lesione da cui parte l'epilessia. Gli effetti negativi delle convulsioni possono essere ridotti monitorando continuamente l'attività elettrica nel cervello.

Per gli studi di rilevamento delle crisi, i segnali EEG dei pazienti epilettici vengono registrati durante i periodi di crisi e durante i periodi senza crisi (Abrol, 2021).

1.2.3 Predizione di crisi epilettiche

Per quanto concerne la previsione dell'epilessia è bene evidenziare che, per una previsione accurata è necessario esaminare registrazioni EEG di lunga durata. Neurologi esperti analizzano l'epilessia studiando i segnali EEG continui registrati per giorni, settimane o anche mesi, azione che richiede una considerevole quantità di tempo e impegno umano.

L'attività cerebrale dei pazienti con epilessia può essere divisa in quattro fasi fondamentali (figura 1.5):

- **fase pre-ictale:** fase antecedente all'attacco vero e proprio. All'interno di questa fase in alcuni soggetti patologici insorge un fenomeno noto come "aura", caratterizzato da un'alterazione sensoriale anticipatoria della crisi epilettica. La sua durata è molto variabile (da 30 minuti a diverse ore);
- **fase ictale:** fase dell'attacco vero e proprio dove si manifestano i sintomi neurologici;
- **fase post-ictale:** fase postuma all'attacco, che conduce il paziente ad una condizione di normalità;
- **fase interictale:** periodo che intercorre tra le crisi dopo che il paziente si è ripreso dalla fase post-ictale.

La previsione della seizure epilettica (ES), quindi, si riferisce all'identificazione dei cambiamenti preictali che possono verificarsi prima delle crisi stesse (Cherian & Kanaga 2022).

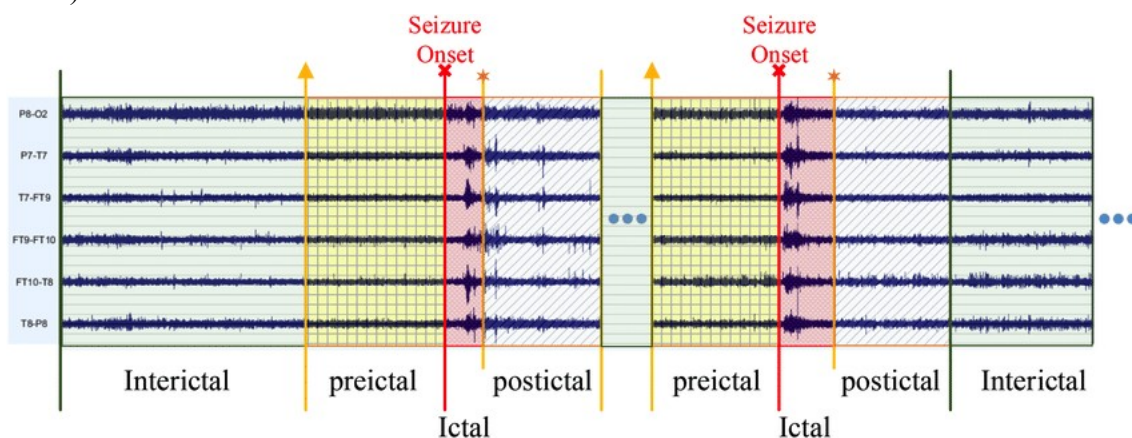


Figura 1.5 Fasi crisi epilettica

Gli studi relativi alla predizione di una crisi epilettica utilizzano modelli di Deep Learning e Machine Learning che, attraverso reti neurali artificiali, permettano di estrarre le caratteristiche dei dati EEG antecedenti la crisi e consentano di predire, con una finestra temporale di lunghezza variabile, un attacco epilettico. Se nel Machine Learning le features devono essere estratte manualmente, nell'ambito del Deep Learning è il modello stesso che apprende le features caratteristiche dei dati di input.

Per farlo, indipendentemente dalle procedure di preprocessing utilizzato, il segnale viene trasformato in spettrogrammi attraverso l'utilizzo della Short Time Fourier Transform (STFT) e passato attraverso la rete neurale che emette un segnale di allarme nel momento in cui sussiste la possibilità di una crisi.

Per la valutazione delle prestazioni del modello si utilizzano due indici: "seizure prediction horizon" (SPH) e "seizure occurrence period" (SOP).

SOP indica la durata, in termini temporali, in cui sussiste la possibilità di una seizure; SPH indica il divario temporale tra l'allarme generato dal modello predittore e l'inizio della SOP. (Truong et al., 2017)

2. DEEP LEARNING

Il Deep Learning (DL) è una branca del Machine Learning, ovvero quel campo scientifico che studia come costruire programmi in grado di apprendere e migliorare le proprie prestazioni relative ad un dominio specifico sulla base dell'esperienza e, infine, generalizzare a nuove situazioni.

Il fulcro di questa tipologia di apprendimento consiste nel creare modelli su più livelli in grado di apprendere delle caratteristiche da un set di dati (training set) e generalizzarle su un nuovo dataset (test set).

Attraverso questa architettura gerarchica multistrato è stato possibile generare modelli da grandi raccolte di dati in modo più preciso. Fortunatamente, la disponibilità di grandi insiemi di dati e il recente sviluppo di unità di elaborazione grafica (GPU) hanno consentito di poter studiare e sviluppare questo tipo di architetture (Cherian & Kanaga 2022).

Nello specifico il Deep Learning è mediato dallo sviluppo delle Reti Neurali Artificiali (ANN) che traggono ispirazione dalle reti neurali biologiche presenti nel nostro cervello. L'ANN è costituita da un insieme di unità connesse chiamate nodi o neuroni. Proprio come il neurone biologico che si trova nel cervello, questi neuroni integrano i segnali di ingresso e li trasmettono ad altri collegati attraverso dei valori matematici chiamati pesi (W) (Acharya et al., 2017)

Più neuroni costituiscono uno strato chiamato layer che può essere suddiviso in 3 gruppi a seconda della sua posizione all'interno della rete:

- input layer, o strato di ingresso;
- hidden layers, o strati nascosti;
- output layer, o strato di uscita.

Come si evince dalla figura 2.1, i modelli Deep Neural Network sono composti da neuroni di input che trasmettono le informazioni ai neuroni nascosti (Hidden Layers) e da qui ai neuroni di output (Goodfellow et al., 2016).

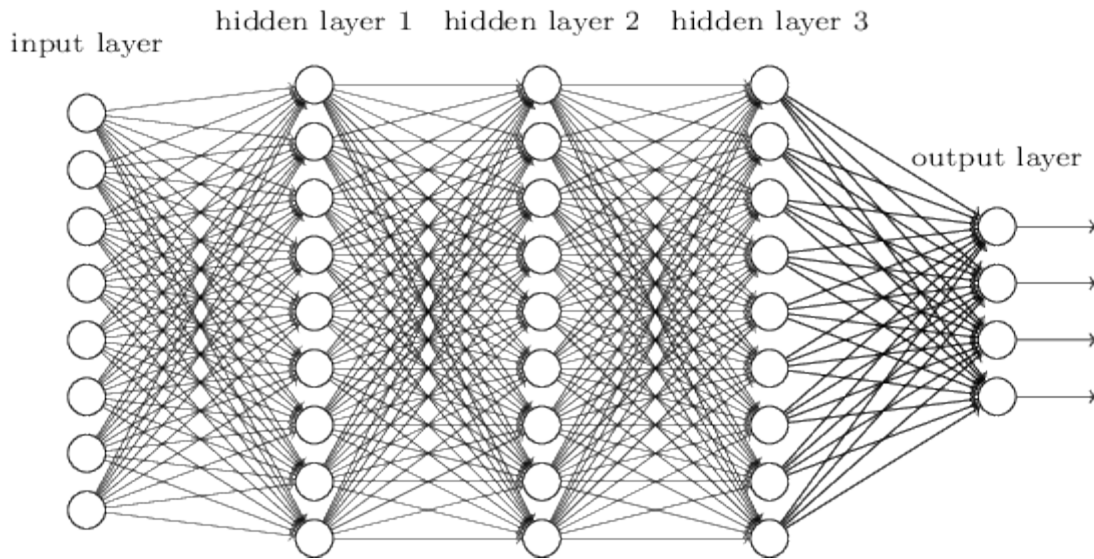


Figura 2.1 Rete neurale feed-forward

Gli strati di neuroni artificiali applicano inizialmente una trasformazione lineare (somma pesata) ai dati ricevuti in input, mentre l'output viene successivamente attivato tramite una funzione di attivazione solitamente non lineare. Le funzioni di attivazione sono espressioni matematiche che definiscono l'output di una rete neurale in quanto determinano se il neurone in questione si attiverà o meno in base ai segnali provenienti dai dati di input o da altri livelli precedenti. La funzione di attivazione è utile anche per normalizzare l'output dei neuroni, costringendolo a essere positivo, negativo o in un certo intervallo (ad esempio $[-1, +1]$) a seconda della scelta di tale funzione.

Per i neuroni di input si utilizza soprattutto la funzione identità, mentre per gli Hidden layers si utilizza: Sigmoide, utile al fine di normalizzare il valore di output di ciascun neurone; Tanh, può portare alla saturazione come la Sigmoide, ma a differenza di quest'ultima il suo output è centrato sullo zero; ReLU, non porta a saturazione permettendo una discesa del gradiente più veloce; Leaky ReLU, consente di applicare l'algoritmo backpropagation anche per valori di input negativi.

Per i neuroni di output si utilizza una funzione Lineare soprattutto nel caso del regression task, la Sigmoide per la classificazione binaria, mentre la Softmax per la classificazione multi-classe (vedi sotto).

È bene evidenziare inoltre che nell'ambito del Machine Learning, e del Deep Learning di conseguenza, ci sono tre framework principali: apprendimento supervisionato, apprendimento non supervisionato ed apprendimento per rinforzo (Goodfellow et al., 2016).

Nell'apprendimento supervisionato, un insegnante esterno mostra al modello come comportarsi fornendo una serie di esempi già correttamente etichettati.

L'output prodotto viene quindi confrontato con quello del supervisore e i parametri vengono modificati al fine di minimizzare la discrepanza tra questi.

Il processo di apprendimento supervisionato nelle reti neurali si può dividere in 3 fasi principali:

- attraverso l'input layer si fornisce alla rete il pattern dei dati di addestramento per ottenere un output;
- si calcola quindi l'errore (attraverso una funzione d'errore) tra output prodotto ed output desiderato;
- l'errore viene poi riportato in ingresso dove si calcola la sua derivata (in funzione) rispetto a ciascun peso nella rete; si aggiorna il modello .

Quindi l'idea è di modificare i parametri θ lungo la direzione che minimizza il gradiente della funzione d'errore.

Dopo aver ripetuto questo processo più volte, è possibile classificare l'output prodotto. Sebbene biologicamente poco plausibili i metodi di apprendimento supervisionato possono essere molto potenti ed efficienti quindi, nel caso in cui un supervisore sia costantemente disponibile, è utile sfruttare sempre questo tipo di apprendimento.

I due task principali nell'ambito dell'apprendimento supervisionato sono classificazione e regressione.

Il processo di classificazione permette di sviluppare un modello (classificatore) che deve categorizzare i dati in due o più classi mentre il processo di regressione produce un output che ha un dominio continuo e non discreto. Nel primo caso, quindi, la rete neurale deve classificare il target entro alcune classi; nel secondo caso, invece, si prevedono i valori di una variabile numerica partendo da una o più variabili.

D'altro canto, nell'apprendimento non supervisionato, non ci sono etichette a priori corrette per i dati; il compito, quindi, consiste nell'estrarre le regolarità statistiche ed usarle per costruire una rappresentazione interna del mondo. Sebbene non sia chiaro come

decidere a priori quali caratteristiche dovrebbero essere apprese dai dati risulta tuttavia abbastanza plausibile a livello biologico.

Nell'apprendimento di rinforzo, l'agente può interagire attivamente con l'ambiente e riceve ricompense e punizioni in risposta alle sue azioni. È bene evidenziare che gli algoritmi per rinforzo sono computazionalmente i più costosi.

Nell'ambito del deep learning le reti neurali più comunemente utilizzate sono la rete neurale convoluzionale (CNN) e la rete neurale ricorrente (RNN).

Le RNN vengono utilizzate per trovare delle sequenze logiche nei dati di input, dove l'output di ogni hidden layer viene sia passato al livello successivo, sia restituito a sé stesso: l'output attuale è un'esperienza combinata del momento presente e del passato.

D'altra parte la CNN, attraverso filtri, estrae le caratteristiche dai dati di input multidimensionali rendendo questa rete neurale ampiamente utilizzata con i dati EEG perché può essere molto efficace per ridurre il rumore (Rasheed et al., 2020).

2.1 Dataset e iperparametri

Per sviluppare queste tipologie di apprendimento la base di partenza è sempre il dataset che verrà poi adattato alla rete neurale.

Viene diviso essenzialmente in 3 parti al fine di implementare l'apprendimento della rete:

- training set: è l'insieme di esempi (o immagini) utilizzati per l'addestramento della neural network e per modificare i pesi (W);
- validation set: è l'insieme di pattern utilizzati per ottimizzare gli iperparametri del modello (learning rate, batch size, weight decay ecc.) e, quindi, trovare il modello migliore;
- test set: è l'insieme di dati utilizzati al fine di valutare la performance finale della rete.

Il dataset è suddiviso in training, validation e test set come mostrato in figura 2.2.

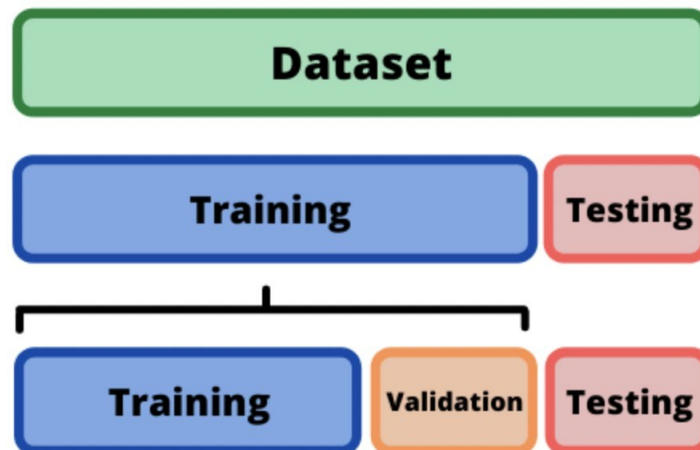


Figura 2.2 Ripartizione del dataset

Per massimizzare il numero di esempi di training si può utilizzare una tecnica chiamata *k-fold cross validation* che consiste nella divisione del dataset in k parti uguali e ad ogni ciclo di cross validazione la *k-esima* parte viene esclusa dal training ed utilizzata come test. La performance finale viene data dalla media dei risultati ottenuti ad ogni ciclo (quindi tra tutte le k ripartizioni).

A seconda delle prestazioni sul set di dati di training/test, possiamo sapere se l'apprendimento ha avuto successo o meno.

Idealmente dovremmo avere un piccolo errore di addestramento, altrimenti il modello non riesce a catturare le caratteristiche dei dati (underfitting) e un errore del test set simile a quello del set di dati di addestramento. Se quest'ultimo è significativamente più alto significa che il modello ha appreso troppo dettagliatamente le caratteristiche del dataset di training e non è in grado di generalizzare a quei dati che sono ancora sconosciuti (overfitting).

Come precedentemente accennato, nel caso dell'apprendimento supervisionato l'obiettivo è diminuire la discrepanza tra output prodotto e target. Questo in termini matematici assume il significato di minimizzare la funzione di costo attraverso la discesa del suo gradiente.

Il learning rate è un iperparametro che determina la dimensione del passo per ogni iterazione verso il minimo della funzione di costo e rappresenta, metaforicamente, la velocità con cui un modello di apprendimento automatico "impara".

Nell'impostare un tasso di apprendimento ci dev'essere un compromesso nella dimensione di quest'ultimo: un learning rate troppo elevato potrebbe evitare di condurre verso un minimo assoluto, mentre un learning rate troppo basso rischierebbe di bloccare l'apprendimento in un minimo locale.

Il batch size, d'altro canto, rappresenta il numero di pattern del training set utilizzato nella stima del gradiente di errore.

Inoltre il weight decay è un iperparametro che consente di aggiungere un termine di penalità ai pesi del modello in modo tale da penalizzarne la complessità e ridurre così il rischio di overfitting.

In ultimo, il Global Average Pooling permette di evitare di aggiungere livelli completamente connessi in seguito ai layer di convoluzione, prendendo la media di ciascuna feature maps. Genera quindi una mappa delle caratteristiche per ciascuna categoria di classificazione nell'ultimo livello convoluzionale ed il risultante vettore viene trasmesso al layer finale. Il vantaggio consiste nell'aver meno parametri da ottimizzare e, quindi, meno rischio di overfitting.

2.2 Reti neurali convoluzionali

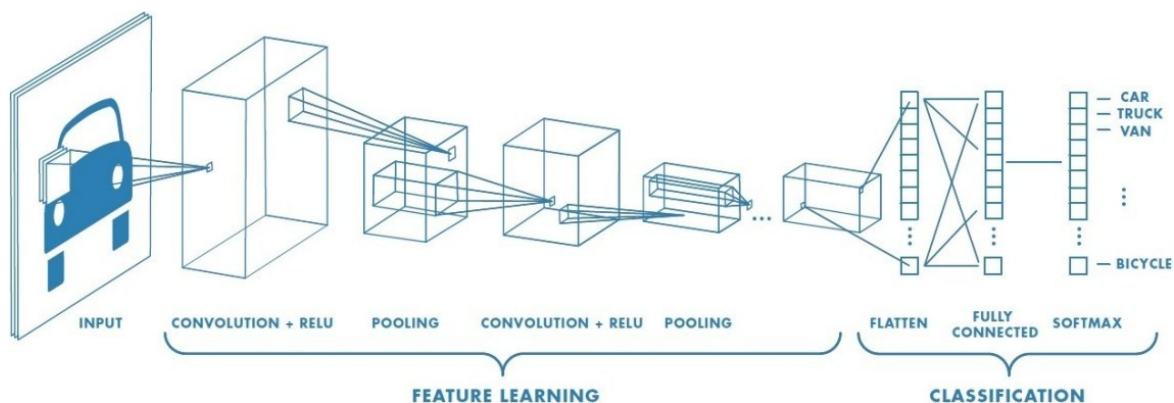


Figura 2.3 Rete convoluzionale

Le Reti Neurali Convoluzionali (CNN) nascono nel 1990 dalla ricerca di Yann LeCun insieme al suo team basandosi sul funzionamento della corteccia visiva del cervello umano (LeCun et al., 1989).

Grazie alle ottime prestazioni ricavate soprattutto in ambito del riconoscimento di immagini, ancora oggi le CNN sono considerate lo “stato dell’arte” per quanto riguarda il riconoscimento di pattern ed immagini.

Le CNN sono reti neurali organizzate in più livelli: i primi strati sono considerati “low-level features extractors”, mentre gli ultimi strati sono considerati “high-level features extractors”.

Nelle immagini le “low-level features” rappresentano, ad esempio, i bordi o i blob (blocchi elementari) che vengono rielaborati per formare delle “high-level features” le quali, attraverso gli ultimi strati della rete, saranno in grado di creare, ad esempio, i contorni di case, cani, gatti o qualsiasi dato fosse presente nell’immagine originale. Le CNN elaborano delle “feature maps” dove ogni elemento corrisponde a dei pixel nell’immagine originale.

Per ottenere questo risultato è necessario effettuare un’operazione detta convoluzione (figura 2.4) (Goodfellow et al., 2016).

Ogni neurone nascosto ha un campo ricettivo locale che codifica per una caratteristica specifica (kernel o filtro). Il numero di neuroni nascosti (ognuno con il proprio peso) definisce quante caratteristiche (kernel) saranno rappresentate ad ogni livello di elaborazione. Tuttavia, ogni filtro svolge il processo di convoluzione con l'intera immagine di input. Pertanto, l'immagine non è più compressa in un vettore a una dimensione, ma in un'immagine bidimensionale filtrata.

I parametri per ogni strato convoluzionale sono:

- numero di "neuroni nascosti": definisce quanti kernel vengono utilizzati ad ogni livello
- dimensione del kernel: definisce il campo ricettivo del kernel
- stride: dimensione del passo di cui si muove il kernel di convoluzione
- padding: strati di input (solitamente a valore zero) che circondano l'immagine, preservando così le dimensioni dell'immagine nella mappa delle caratteristiche.

Il processo di convoluzione viene svolto a più livelli e permette, attraverso l'uso di filtri, di estrarre delle caratteristiche (o features) delle immagini di cui si vuole analizzare il contenuto.

Come mostrato in figura 2.4, tali filtri vengono fatti scorrere sulle diverse posizioni dell'immagine in input: per ogni posizione viene generato un valore di output, eseguendo il prodotto scalare tra il filtro e la porzione dell'input corrispondente (campo ricettivo).

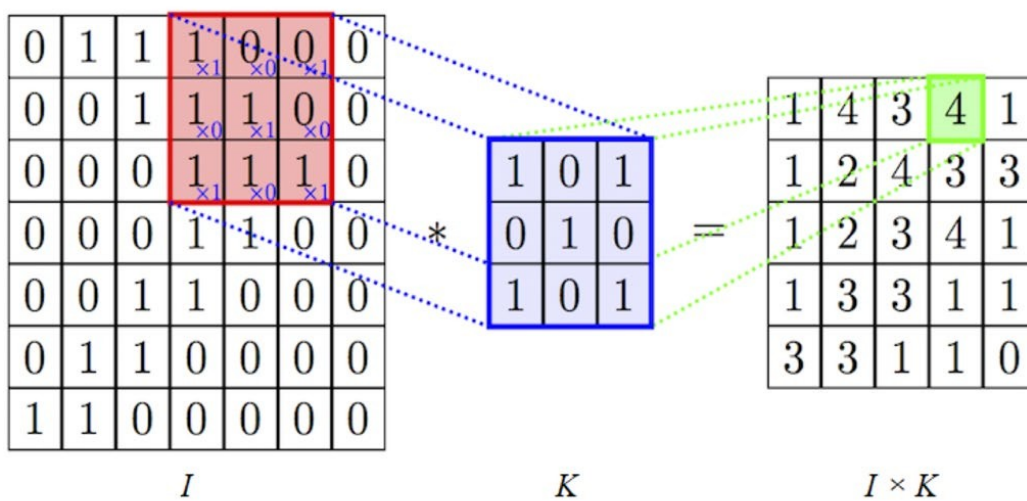


Figura 2.4 Processo di convoluzione

A differenza di una rete feed-forward tradizionale che lavora “sull’informazione generale dell’immagine” una CNN lavora e classifica l’immagine basandosi su particolari caratteristiche della stessa. In altre parole, a seconda del tipo di filtro utilizzato è possibile identificare sull’immagine di riferimento elementi diversi, ad esempio i contorni delle figure, le linee verticali, le linee orizzontali, le diagonali, etc.

Tipicamente ogni layer convoluzionale viene fatto seguire da uno di Max-Pooling, il quale riduce la dimensione della matrice, ma aumenta di conseguenza il livello di “astrazione”. Si passa quindi da filtri elementari, come appunto linee verticali e orizzontali, a filtri via via più sofisticati, in grado di riconoscere features sempre più complesse.

I livelli convoluzionali vengono successivamente fatti seguire da uno o più livelli lineari fully-connected che permettono la classificazione delle immagini.

I modelli convoluzionali si collocano a formare un ponte tra le neuroscienze e l’intelligenza artificiale. Infatti, come accennato precedentemente, attraverso un’interazione con l’informatica e l’ingegneria le CNN sono un esempio di architetture che si ispirano al cervello e, in particolare, al sistema visivo.

Durante il processing di un’immagine da parte della CNN le caratteristiche più complesse vengono apprese a livello degli strati convoluzionali superiori. Questo fenomeno è in analogia con quanto avviene nella via ventrale.

Ogni stazione di elaborazione e via di connessione nel sistema visivo contiene un insieme unico di neuroni con caratteristiche e funzioni diversificate.

Le prime tappe dell’elaborazione delle informazioni visive avvengono nella retina, la quale, situata nella parte posteriore dell’occhio, è composta da coni e bastoncelli, recettori che trasformano l’energia luminosa in attività neuronale (Bear et al., 2016)

In seguito l’informazione passa al nucleo genicolato laterale (NGL) attraverso il nervo e tratto ottico. Gli assoni delle cellule postsinaptiche del nucleo genicolato laterale formano le radiazioni ottiche che terminano nella corteccia visiva primaria (detta area visiva 1, o area V1). L’elaborazione delle informazioni circa posizione e forma di oggetti statici o in movimento nel campo visivo è una funzione altamente specializzata dell’Area V1.

Da essa si diramano due sottosistemi che costituiscono la via visiva ventrale e la via visiva dorsale (Figura 2.5).

La via ventrale, o occipito-temporale, collega la corteccia visiva primaria (V1) alle porzioni inferiori del lobo temporale ed è importante per l'identificazione percettiva degli oggetti, il riconoscimento visivo di forme, facce e parole.

La via occipito-parietale, nota anche come via dorsale, è collegata al lobo parietale ed è fondamentale per le attività spaziali e visuo-motorie, compreso il rilevamento della posizione e del movimento degli oggetti.

La corteccia temporale inferiore è raggiunta dalla via ventrale, che parte dalla corteccia V1 e viaggia attraverso le aree visive V2 e V4.

L'area V2 riceve connessioni da V1 ed invia segnali verso V3, V4, V5. Molti dei neuroni di quest'area sono regolati da caratteristiche visive semplici (orientamento, frequenza spaziale, dimensioni, colore, forma) ma anche più complesse.

L'area V4 è specializzata soprattutto nel riconoscimento dei colori, infatti, un danno bilaterale in questa zona causa acromatopsia. Infine, nell'area IT convergono informazioni sia visive che mnestiche; i suoi neuroni hanno campi recettivi molto ampi che rispondono indipendentemente dalle dimensioni dell'oggetto, dal suo orientamento e dalla sua posizione nel campo visivo. Una parte di quest'area si estende lungo il solco fusiforme, quindi una lesione in questa zona causa prosopagnosia (Bear et al., 2016).

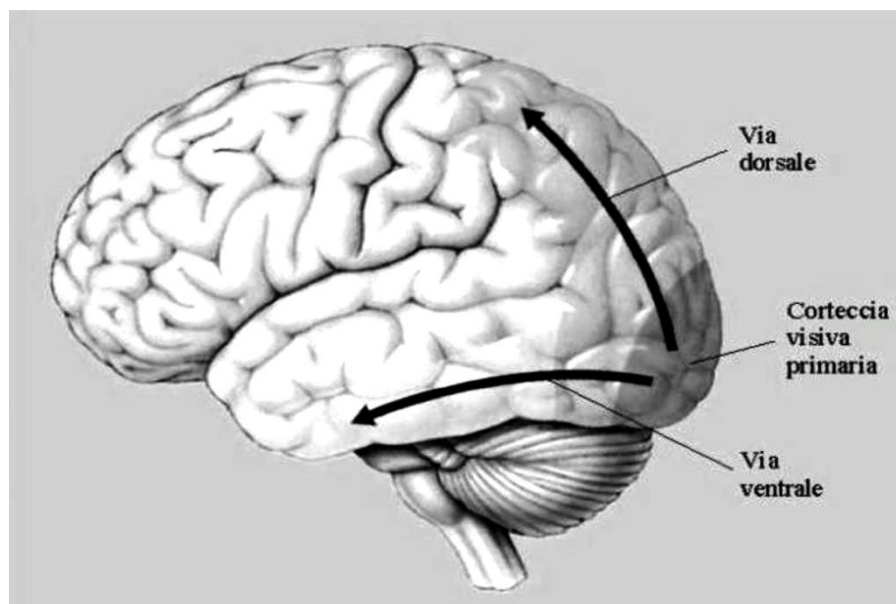


Figura 2.5 Via visiva dorsale ventrale

Arrivati a questo punto appare chiaro come vi sia un trend crescente in termini di complessità di analisi delle caratteristiche visive degli oggetti a mano a mano che ci si inoltra all'interno del sistema visivo. Basti pensare, ad esempio, come nella corteccia visiva primaria l'informazione principale riguardi l'identificazione di bordi, orientamento di figure ecc, mentre a livello temporale inferiore (Fusiform Face Area) si possano riconoscere i volti.

Le reti convoluzionali traggono ispirazione dal sistema visivo, i layers sono considerati come i neuroni nelle regioni cerebrali associate all'elaborazione di stimoli sensoriali visivi (Rasheed et al., 2020). Ai livelli inferiori i filtri saranno in grado di elaborare semplici informazioni (linee, bordi ecc) mentre negli strati convoluzionali superiori la rete sarà in grado di elaborare features di livello molto più complesso tra cui volti, macchine, case ecc (Goodfellow et al., 2016).

Come citato nei capitoli precedenti questa specializzazione è resa possibile attraverso un livello di Pooling che diminuisce la dimensione della matrice aumentandone il livello di astrazione. Ciascun layer convoluzionale può quindi essere considerato come l'approssimazione di una singola area visiva (lungo la via ventrale) come V1, V2, V4 e IT. Questo processo consente ai campi ricettivi dei singoli neuroni di aumentare la propria dimensionalità procedendo verso gli ultimi strati della rete e le features a cui rispondono diventano via via più complesse.

Le CNN sono nate grazie agli sviluppi in campo medico-neuroscientifico a partire dalla seconda metà del Novecento. Hubel e Wiesel hanno scoperto due principali tipologie cellulari nella corteccia visiva primaria (V1) dei gatti (Hubel & Wiesel, 1962).

Il primo tipo, le celle semplici, rispondono a barre di luce poste in posizioni spaziali specifiche. Ogni cellula, infatti, si attiva maggiormente a seconda dell'orientamento spaziale che la barra assume, mentre, la sua risposta, diminuisce quando l'angolo della barra cambia rispetto all'orientamento preferenziale. Il secondo tipo, le cellule complesse, hanno ancora orientamenti preferiti ma possono rispondere altrettanto fortemente a una barra in diverse posizioni. Hubel e Wiesel hanno concluso che quest'ultime ricevono input da diverse cellule semplici, tutte con lo stesso orientamento ma con posizioni preferite leggermente diverse (Lindsay, 2020).

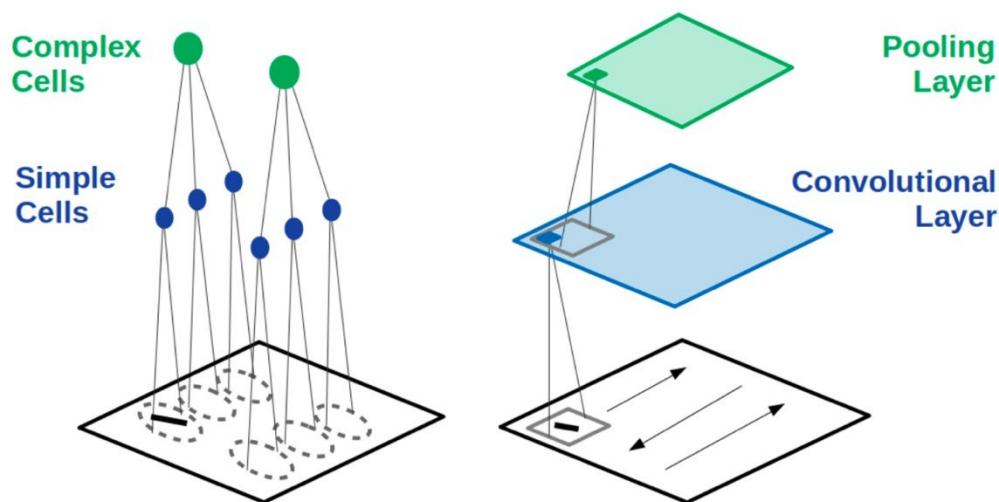


Figura 2.6 Modello Hubel e Wiesel comparato alla CNN

Partendo da questi risultati, nel 1980 Fukushima ha trasformato le scoperte di Hubel e Wiesel in un modello precursore delle reti convoluzionali (Neocognitron) (Dissanayake et al., 2020).

Esso contiene due tipi cellulari principali: le cellule S, che prendono il nome dalle cellule semplici, replicano le loro caratteristiche di base formando più layer che condividono le stesse caratteristiche visive. In particolare, una griglia 2D di pesi viene applicata in ciascuna posizione nell'immagine di input per creare le risposte delle cellule S. La risposta delle cellule C (che prendono il nome dalle cellule complesse) è una funzione non lineare di diverse cellule S provenienti dallo stesso layer ma in posizioni diverse. Dopo uno strato di cellule semplici e complesse (metaforicamente rappresentanti V1), il Neocognitron ripete semplicemente di nuovo il processo: l'output del primo strato di cellule complesse

funge da input per il secondo strato di cellule semplici e così via. Ripetendo questo processo si crea un modello gerarchico che imita la via visiva ventrale nel suo insieme.

Quindi, la rete neurale convoluzionale proviene dal campo della visione artificiale, ma l'ispirazione dal lavoro di Hubel e Wiesel è chiaramente visibile in essa (Figura 2.6). Le moderne CNN iniziano adattando una serie di filtri all'immagine di input creando "feature maps" simili ai layer delle cellule S nel Neocognitron. Viene quindi applicato il livello di pooling, che genera un output simile a quello delle cellule C. Alla fine del processo vengono aggiunti i fully-connected layers e l'ultimo livello contiene tanti neuroni quante sono le categorie

2.3 Transfer Learning

Il transfer learning è un processo che consiste nell'utilizzo di una rete neurale sulla quale è stato già svolto un training attraverso un set di dati di grandi dimensioni.

Con questo metodo un modello pre-allenato per svolgere una precisa attività viene riutilizzato come punto di partenza per lo sviluppo di un modello destinato all'esecuzione di un problema simile a quello per cui è stato progettato, mediante l'utilizzo di uno o più livelli dello stesso.

I modelli gerarchici, come quelli di Deep Learning, facilitano la condivisione della rappresentazione: invece di apprendere ogni compito da zero, semplicemente perfezionano le caratteristiche di alto livello che caratterizzano ogni compito.

La logica alla base del transfer learning è che la rappresentazione condivisa tra i compiti può supportare meglio l'inferenza e la generalizzazione, a condizione che i compiti siano in qualche modo correlati.

Tale sistema, tipicamente, viene impiegato all'interno di un contesto di apprendimento supervisionato permettendo di riutilizzare il modello pre-addestrato nella sua interezza (aggiungendo al modello un nuovo classificatore) o come base per la creazione di un altro modello di rete neurale per lo svolgimento di un determinato task.

L'intuizione alla base del transfer learning è che se un modello viene addestrato su un set di dati sufficientemente ampio e generale, lo stesso fungerà efficacemente da modello generico.

Uno dei maggiori vantaggi del transfer learning consiste nel limitare l'addestramento e l'elaborazione ad un numero inferiore di parametri. Infatti, sbloccando alcuni dei livelli superiori del modello pre-allenato, non solo si riduce il numero di parametri, ma è possibile un addestramento ulteriore e maggiormente specifico del modello perfezionando le feature maps di livello superiore.

L'architettura risultante è simile a quella di figura 2.7.

Si può, quindi, implementare un sistema di visione artificiale e riciclare le conoscenze visive in modo tale da addestrare solo gli ultimi strati fully-connected.

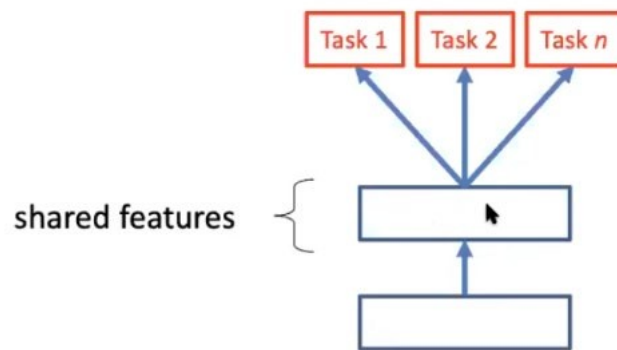


Figura 2.7 Architettura Transfer Learning

La rete VGG 16 è un modello pre-allenato che mediante procedure di transfer learning è frequentemente utilizzato per task di classificazione e object detection.

Il training di questa rete è avvenuto su un dataset comprendente circa 1.2 milioni di immagini, a cui si sommano altre 100.000 immagini per la fase di testing.

Il modello è basato su una rete convoluzionale e comprende 16 layers di elaborazione 3x3 sovrapposti con profondità crescente e 5 livelli di MaxPooling utili per ridurre la dimensionalità della matrice. Infine, due fully-connected layers premettono la classificazione.

3. METODOLOGIA

Lo scopo di questa tesi è la predizione di un attacco epilettico tramite l'analisi del tracciato elettroencefalografico attraverso una rete neurale artificiale.

Per la realizzazione di questo progetto abbiamo collaborato con l'ospedale di Conegliano, i cui medici ci hanno fornito svariate ore di registrazioni elettroencefalografiche provenienti da 20 pazienti con diagnosi di Epilessia.

I file contenenti le registrazioni EEG sono in formato "TRC" e da questi, attraverso l'utilizzo del software Brainstorm (Brainstorm è un'applicazione open source dedicata all'analisi delle registrazioni cerebrali: MEG, EEG, fNIRS, ECoG, elettrodi di profondità ed elettrofisiologia multiunità), abbiamo estratto i corrispondenti file in formato "txt" contenenti le informazioni temporali circa l'inizio e la fine di ogni seizure epilettica. Quest'ultimi sono stati utili per la costruzione delle "label" del dataset.

Le registrazioni EEG dei 20 pazienti non sono state ottenute tutte con lo stesso numero di canali per ciascun soggetto; quindi, abbiamo selezionato 16 canali in comune tra tutti i pazienti per comporre il dataset.

In accordo con le ricerche svolte da Ohiraa e colleghi (2018) abbiamo rimosso dal dataset l'ora successiva alla crisi epilettica in quanto, nel periodo post-ictale, il rumore causato dalla fase ictale può creare delle distorsioni nei dati e, di conseguenza, una diminuzione in termini di accuratezza.

Per quanto riguarda il periodo pre-ictale, la sua durata varia di paziente in paziente; quindi, la scelta della finestra temporale antecedente all'attacco epilettico è una "design choice". Per non ridurre le prestazioni del modello con una finestra troppo ampia o per evitare di dare poco preavviso al paziente, nel nostro progetto abbiamo selezionato le 2 ore antecedenti la crisi.

Considerando un'ora di tempo per il post-ictale e due ore di periodo pre-ictale, nel momento in cui un paziente ha manifestato due crisi consecutive in un intervallo di tempo inferiore a 3 ore, la seconda crisi è stata rimossa dal dataset (Ohiraa et al., 2018).

Le operazioni di preprocessing e processing sono state sviluppate attraverso Python, un linguaggio di programmazione di alto livello orientato a "oggetti" che può vantare di

essere uno de linguaggi per eccellenza nello sviluppo di reti neurali artificiali ed in generale di modelli per l'analisi di dati avanzata.

Il “seizure prediction process” è un task di apprendimento supervisionato che richiede che i dati siano etichettati (label), quindi suddivisi in categorie.

Con l'utilizzo di Python abbiamo suddiviso il segnale preictale in 4 categorie, ciascuna dell'ampiezza di 1800 secondi (30 minuti).

Al segnale EEG proveniente da ciascun canale è stato effettuato un downsampling a 256 Hz che permette, quindi, di ridurre il numero di campioni utilizzati pur mantenendo il tracciato informativo.

In seguito, attraverso la funzione “spectrogram” proveniente dalla libreria “scipy” abbiamo creato ogni 5 secondi gli spettrogrammi relativi a ciascun canale. Questa funzione applica al segnale la Short-Time Fourier Transform (STFT) la quale, a differenza della trasformata di Fourier dove il segnale è rappresentato sotto forma di Hz in funzione della potenza, divide il segnale temporale totale in segmenti brevi e, quindi, per ogni segmento più corto, ne calcola la trasformata di Fourier. Il risultato finale consente quindi di visualizzare gli spettri che cambiano in funzione del tempo: come mostrato in figura 3.1 nelle ascisse avremo l'informazione relativa al tempo, nelle ordinate quella relativa alla frequenza.

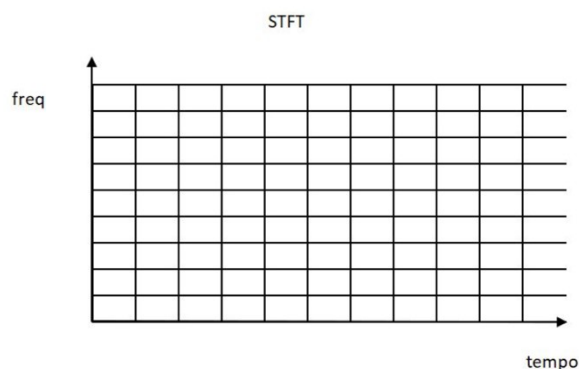


Figura 3.1 Short-Time Fourier Transform

Una volta utilizzata questa tecnica abbiamo opportunamente filtrato il segnale utilizzando le apposite funzioni dalla libreria “mne”.

Un filtro passa-alto di 1 Hz è stato utile al fine di rimuovere le fluttuazioni di base e gli artefatti da movimento mentre un filtro passa-basso di 125 Hz è stato applicato al tracciato al fine di evitare il fenomeno dell'aliasing.

Per eliminare il rumore di rete abbiamo utilizzato un notch filter a 50 Hz e, siccome questo filtro è spesso seguito anche da filtri nelle sue armoniche, abbiamo applicato la stessa tipologia di filtro anche a 100 Hz in modo da evitare che il rumore di rete si ripercuota anche nelle altre sue bande, specialmente perché il segnale è registrato in ambiente non schermato.

A questo punto il dataset è pronto. I dati risultano essere molto complessi e difficili da analizzare rispetto, ad esempio, all'immagine di un oggetto; quindi, il 90% degli spettrogrammi hanno composto il training set, mentre i restanti il test set.

Il dataset è stato quindi normalizzato utilizzando la media e la deviazione standard del training set e, infine, passato attraverso le reti neurali.

I modelli sono stati sviluppati grazie alla libreria Keras, un framework open source specifico per la costruzione di modelli di Deep Learning; Keras permette la creazione di reti artificiali in modo intuitivo fornendo tutte le più comuni implementazioni del modelli DL.

Per la predizione dell'attacco epilettico abbiamo testato due reti neurali: VGG-16 (Transfer learning) ed una "Vanilla CNN".

3.1 VGG 16

Come accennato nel capitolo precedente la rete neurale VGG 16 è un tipo di rete convoluzionale composta da 16 layers pre-addestrati nel riconoscimento di oggetti attraverso il dataset ImageNet Large Scale Visual Recognition Challenge (ILSVRC). In questo modello ci sono 13 strati convoluzionali, cinque strati di Max Pooling e tre fully-connected layers dove, però, solo 16 strati di questi possono apprendere, quindi essere addestrati al fine di modificarne i pesi.

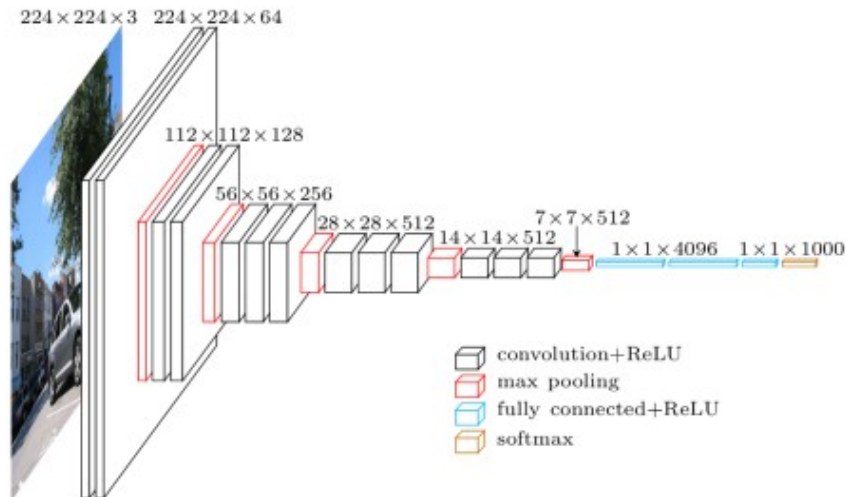


Figura 3.2 Architettura VGG 16

Come evidenziato dalla figura 3.2, il layer Conv-1 ha 64 filtri, Conv-2 ha 128 filtri, Conv-3 ha 256 filtri, Conv-4 e Conv-5 hanno 512 filtri. Dopo i livelli convoluzionali ci sono tre livelli completamente connessi (FC), i primi due dei quali hanno ciascuno 4096 canali, mentre il terzo esegue la classificazione ILSVRC a 1000 categorie e quindi ha 1000 canali (uno per ogni classe). Lo strato finale è lo strato softmax.

Al fine di utilizzare la procedura denominata Transfer Learning abbiamo rimosso gli ultimi tre layers completamente connessi (figura 3.3) ponendo come “falso” la possibilità di includere quest’ultimi (include_top=False) e di poter allenare i pesi del modello.

```
base_model = VGG16(input_shape=IMAGE_SIZE + [3], weights="imagenet", include_top=False)
base_model.trainable = False ## Not trainable weights
base_model.summary()
```

Figura 3.3 Operazione di Transfer Learning


```

Model: "vgg16"
-----
Layer (type)                Output Shape                Param #
-----
input_4 (InputLayer)        [(None, 39, 65, 3)]        0
block1_conv1 (Conv2D)        (None, 39, 65, 64)         1792
block1_conv2 (Conv2D)        (None, 39, 65, 64)         36928
block1_pool (MaxPooling2D)   (None, 19, 32, 64)         0
block2_conv1 (Conv2D)        (None, 19, 32, 128)        73856
block2_conv2 (Conv2D)        (None, 19, 32, 128)        147584
block2_pool (MaxPooling2D)   (None, 9, 16, 128)         0
block3_conv1 (Conv2D)        (None, 9, 16, 256)         295168
block3_conv2 (Conv2D)        (None, 9, 16, 256)         590080
block3_conv3 (Conv2D)        (None, 9, 16, 256)         590080
block3_pool (MaxPooling2D)   (None, 4, 8, 256)          0
block4_conv1 (Conv2D)        (None, 4, 8, 512)          1180160
block4_conv2 (Conv2D)        (None, 4, 8, 512)          2359808
block4_conv3 (Conv2D)        (None, 4, 8, 512)          2359808
block4_pool (MaxPooling2D)   (None, 2, 4, 512)          0
block5_conv1 (Conv2D)        (None, 2, 4, 512)          2359808
block5_conv2 (Conv2D)        (None, 2, 4, 512)          2359808
block5_conv3 (Conv2D)        (None, 2, 4, 512)          2359808
block5_pool (MaxPooling2D)   (None, 1, 2, 512)          0
-----
Total params: 14,714,688
Trainable params: 0
Non-trainable params: 14,714,688
-----

```

Figura 3.4 Rimozione ultimi tre livelli completamente connessi VGG 16

Come si evince dalla figura 3.4, gli ultimi tre livelli completamente connessi sono stati rimossi ed il totale dei parametri addestrabili è nullo.

Una volta terminato questo processo sono stati creati 3 fully-connected layers, uniti al termine del modello VGG 16. Questi tre layers avevano rispettivamente 50, 20 e 4 nodi

(corrispondenti alle 4 categorie da classificare) fornendo la possibilità di addestrare un totale di 26 754 parametri (figura 3.5).

Layer (type)	Output Shape	Param #
vgg16 (Functional)	(None, 1, 1, 512)	14714688
flatten_5 (Flatten)	(None, 512)	0
dense_15 (Dense)	(None, 50)	25650
dense_16 (Dense)	(None, 20)	1020
dense_17 (Dense)	(None, 4)	84

Total params: 14,741,442		
Trainable params: 26,754		
Non-trainable params: 14,714,688		

Figura 3.5 Modello finale

Sebbene l'input richiesto fossero immagini RGB 224x224, per evitare di denaturare gli spettrogrammi abbiamo mantenuto il loro formato originale equivalente a 39x62 pixels.

LEARNING RATE	10^{-3}	5×10^{-4}
WEIGHT DECAY	0	10^{-3}

Tabella 3.2 Iperparametri VGG 16

Prima del training finale la rete è stata addestrata su 20 epoche per alcune combinazioni di learning rate e weight decay con l'obiettivo di identificare quella che producesse l'accuratezza più elevata. Per il primo iperparametro sono stati scelti i valori di 10^{-3} e 5×10^{-4} ; per il secondo invece sono stati selezionati valori pari a 0 e 10^{-3} (tabella 3.1). Una volta trovata la miglior combinazione, la rete è stata nuovamente addestrata sul training set per un totale di 100 epoche e testata sul test set con un batch size uguale a 64. Infine è stata nuovamente addestrata su una nuova configurazione con batch size pari a 128 e learning rate uguale a 5×10^{-3} per 300 epoche e testata sul test set.

3.2 Vanilla CNN

Layer (type)	Output Shape	Param #
conv2d_45 (Conv2D)	(None, 61, 38, 128)	18048
max_pooling2d_45 (MaxPooling2D)	(None, 30, 19, 128)	0
conv2d_46 (Conv2D)	(None, 29, 18, 256)	131328
max_pooling2d_46 (MaxPooling2D)	(None, 14, 9, 256)	0
conv2d_47 (Conv2D)	(None, 13, 8, 512)	524800
max_pooling2d_47 (MaxPooling2D)	(None, 6, 4, 512)	0
flatten (Flatten)	(None, 12288)	0
dense_15 (Dense)	(None, 4)	49156

=====
Total params: 723,332
Trainable params: 723,332
Non-trainable params: 0
=====

Figura 3.6 Rete convoluzionale

Come mostrato in figura 3.6 la rete neurale convoluzionale allenata “da zero” è formata da 3 layers convoluzionali ognuno avente la ReLu come funzione di attivazione e seguito da un livello di MaxPooling (2 x 2) per ridurre la dimensionalità della matrice ed aumentare il livello di astrazione delle features.

Il primo livello di convoluzione è formato da 128 neuroni ed un kernel 2 x 2; nel secondo blocco il numero di nodi è 256 mentre la dimensione del kernel è la medesima; nel terzo ed ultimo layer convoluzionale il filtro 2 x 2 viene affiancato dalla presenza di 512 neuroni artificiali.

Viene svolta quindi un’operazione di “Flatten” che permette di generare un vettore contenente le informazioni risultanti dal processo di convoluzione e, in seguito, mediante un ultimo vettore che utilizza la Softmax come funzione di attivazione, produrre un output di dimensione pari a 4, ovvero il numero di categorie nelle quali è stato suddiviso il segnale preictale.

La Softmax è utilizzata al fine di calcolare la probabilità di appartenenza a ciascuna classe con un output nel range (0, 1); alla classe target è attribuita la probabilità più elevata.

Prima di iniziare il training della rete abbiamo adottato una procedura di k-fold cross validation per massimizzare il numero di esempi di training, con un numero di fold pari a 3 ($k=3$).

La funzione di perdita (Loss function) è la Sparse Categorical Cross Entropy, che permette di calcolare la discrepanza tra output prodotto e target e richiede un numero di neuroni di output pari al numero di categorie entro cui classificare l'input.

GLOBAL AVERAGE POOLING	True		False		
LEARNING RATE	10^{-4}	5×10^{-4}	10^{-3}	5×10^{-3}	10^{-2}
BATCH SIZE	32		64		256

Tabella 3.2 Iperparametri del modello

Mediante una procedura di “grid search” abbiamo addestrato la rete su 100 epoche per ogni combinazione allo scopo di trovare quella che producesse l'accuratezza più elevata (tabella 3.2).

La rete è stata quindi nuovamente addestrata sul training set e testata sul test set adottando come iperparametri quelli che hanno fornito l'accuratezza migliore.

La matrice di confusione e le “heat maps” sono state prodotte per permettere l'interpretazione dei risultati ottenuti.

4. RISULTATI

4.1 VGG 16

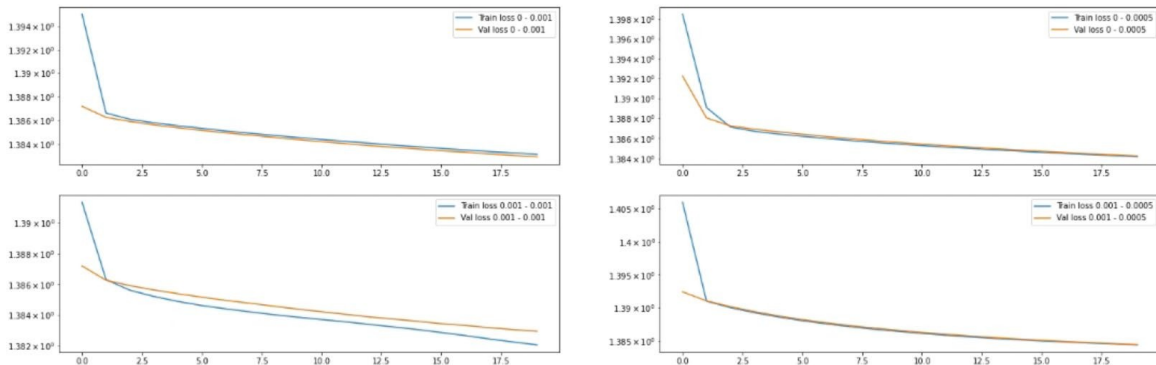


Figura 4.3 Risultati della loss del processo di hyperparameters tuning

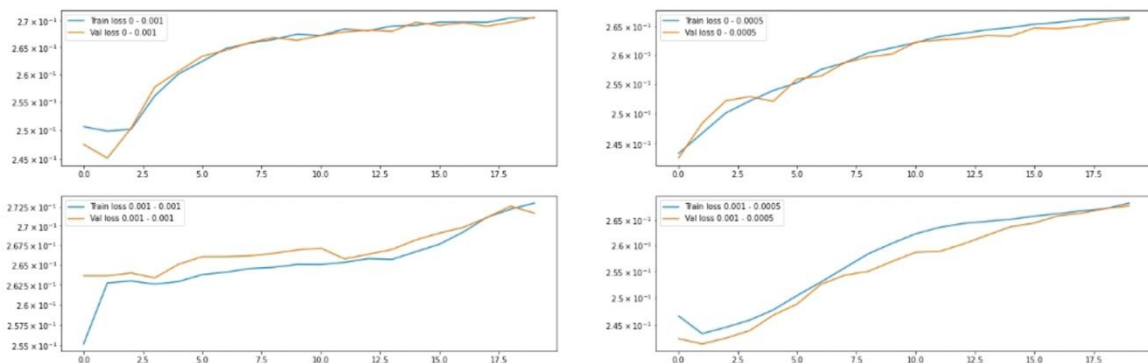


Figura 4.4 Risultati dell'accuratezza del processo di hyperparameters tuning

In figura 4.1 e 4.2 sono visibili i risultati del processo di “hyperparameters tuning”. Indipendentemente dalla combinazione tra i due iperparametri i risultati non sembrano tuttavia molto soddisfacenti.

Esaminando la funzione d'errore, infatti, in tutti i grafici la decrescita sembra essere buona e senza oscillazioni o fenomeni che possano indicare la presenza di overfitting; tuttavia, il range di decrescita è eccessivamente ristretto (~ 0.01), segno che la rete non riesce ad apprendere le features caratteristiche delle immagini in input.

In analogia con questi risultati, a livello grafico l'accuratezza aumenta epoca dopo epoca ma il range di crescita dopo 20 epoche è limitato e compreso tra 0.240 e 0.273.

Sebbene i risultati non siano soddisfacenti l'apprendimento di VGG è stato prolungato fino a 100 epoche utilizzando learning rate e weight decay equivalenti a 10^{-3} .

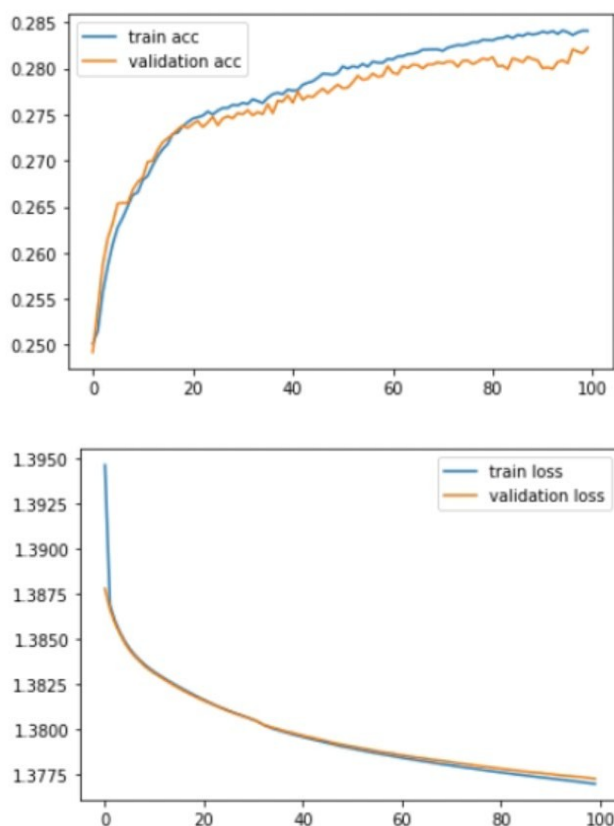


Figura 4.3 Risultati di accuratezza (a) e loss (b) su 100 epoche di training

I risultati in figura 4.3 confermano il trend mostrato con il processo di “hyperparameters tuning”: la decrescita della loss e l'aumento dell'accuratezza a livello grafico non mostrano alcuna imperfezione; tuttavia, il range di miglioramento della performance è troppo limitato e l'apprendimento avviene in maniera estremamente lenta.

Questi risultati mettono in luce un problema di lenta convergenza, dovuto ad una combinazione di iperparametri non ottimale.

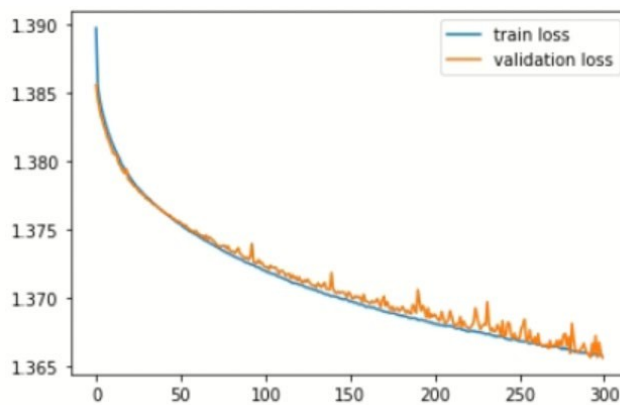
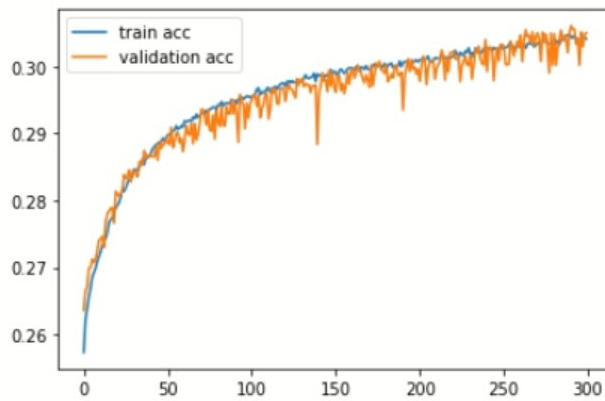


Figura 4.4 Risultati di accuratezza (a) e loss (b) su 300 epoche di training

Inoltre, come mostra la figura 4.4, è stato svolto un training più robusto su 300 epoche utilizzando un learning rate pari a 5×10^{-3} e un valore nullo per il weight decay. Sebbene a livello grafico loss e accuracy sembrano poter migliorare ancora al crescere del numero di epoche, i risultati non mostrano miglioramenti consistenti durante il training, sempre a causa del range limitato di miglioramento della performance.

L'accuratezza in fase di test, infatti, è del 25.12%, indice di un classificatore che categorizza in modo casuale le immagini nelle 4 categorie (Figura 4.5).

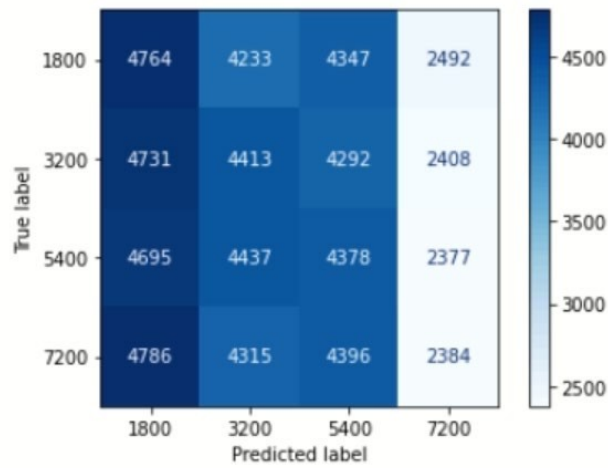


Figura 4.5 Matrice di confusione VGG 16

In conclusione, l'operazione di transfer learning sulla VGG non sembra aver dato i risultati sperati in quanto la rete neurale non è in grado di apprendere a sufficienza i pattern caratteristici degli spettrogrammi dati come input.

4.2 Vanilla CNN

	gap	learning rate	batch size	accuracy	best epoch
0	True	0.0001	32	0.849762	93
1	True	0.0001	64	0.831757	89
2	True	0.0001	256	0.760572	98
3	True	0.0005	32	0.866631	99
4	True	0.0005	64	0.877676	97
5	True	0.0005	256	0.863984	95
6	True	0.0010	32	0.822528	57
7	True	0.0010	64	0.839398	97
8	True	0.0010	256	0.850443	99
9	True	0.0050	32	0.266435	1
10	True	0.0050	64	0.264695	0
11	True	0.0050	256	0.500870	63
12	True	0.0100	32	0.269158	1
13	True	0.0100	64	0.265451	0
14	True	0.0100	256	0.275588	2
15	False	0.0001	32	0.861185	88
16	False	0.0001	64	0.859899	99
17	False	0.0001	256	0.823663	99
18	False	0.0005	32	0.805280	39
19	False	0.0005	64	0.837809	55
20	False	0.0005	256	0.863000	73
21	False	0.0010	32	0.660035	32
22	False	0.0010	64	0.759891	20
23	False	0.0010	256	0.830850	98
24	False	0.0050	32	0.253953	4
25	False	0.0050	64	0.270595	0
26	False	0.0050	256	0.285574	1
27	False	0.0100	32	0.253953	1
28	False	0.0100	64	0.253877	0
29	False	0.0100	256	0.361374	94

Tabella 4.1 Risultati training CNN

I risultati riportati in tabella 4.1 sono stati ricavati dalla tesi di laurea magistrale dello studente del corso di laurea in Physics of Data presso l'Università di Padova che ha collaborato al progetto di ricerca.

Le 30 possibili combinazioni riportate nascono da un processo chiamato "hyperparameters tuning" che permette di mettere in risalto la combinazione dei tre iperparametri (GAP, learning rate, batch size) che producono una migliore accuratezza dei risultati.

Analizzando la tabella in questione emerge da subito come il learning rate abbia l'influenza maggiore nell'accuracy in quanto, quando tale valore diventa maggiore o uguale a 0.005, il valore della corretta predizione del modello diminuisce drasticamente. La dimensione del batch size, invece, non sembra influire.

Infine, la presenza del Global Average Pooling sembra essere un fattore utile all'incremento dell'accuratezza sebbene, confrontando i risultati, non rilevante tanto quanto il learning rate.

	gap	learning rate	batch size	accuracy	best epoch
4	True	0.0005	64	0.877676	97

Tabella 4.2 Migliore combinazione degli iperparametri

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 61, 38, 128)	18048
max_pooling2d (MaxPooling2D)	(None, 30, 19, 128)	0
conv2d_1 (Conv2D)	(None, 29, 18, 256)	131328
max_pooling2d_1 (MaxPooling2D)	(None, 14, 9, 256)	0
conv2d_2 (Conv2D)	(None, 13, 8, 512)	524800
max_pooling2d_2 (MaxPooling2D)	(None, 6, 4, 512)	0
global_average_pooling2d (GlobalAveragePooling2D)	(None, 512)	0
dense (Dense)	(None, 4)	2052
=====		
Total params: 676,228		
Trainable params: 676,228		
Non-trainable params: 0		

Figura 4.6 CNN con layer di GAP

In questo studio la combinazione migliore di iperparametri in termini di accuratezza è data dalla presenza del Global Average Pooling, un tasso di apprendimento pari a 5×10^{-4} ed un valore del batch size uguale a 64 (Tabella 4.2).

L'accuracy ottenuta attraverso questo modello mostrato in figura 4.7 è pari all'87.77%.

Una volta testata la rete con i parametri di riferimento appena citati, l'accuratezza prodotta nel test set è 87.92%.

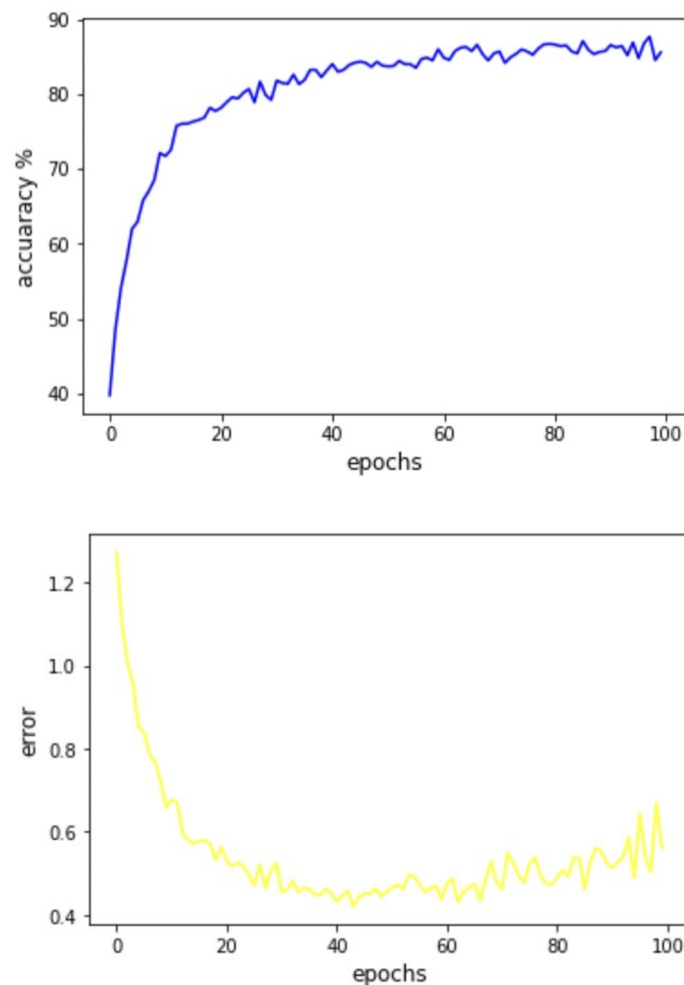


Figura 4.7 Grafici di accuratezza e loss attraverso le 100 epoche

Dal grafico dell'accuratezza emergono risultati soddisfacenti: l'accuratezza cresce in modo graduale nelle prime 40 epoche, mentre nelle successive 60 cresce molto lentamente fino ad arrivare all'epoca 97 dove troviamo il risultato migliore.

Per quanto concerne il grafico della Loss emerge una forte decrescita nelle prime 20 epoche, il valore dell'errore resta costante tra l'epoca 40 e 60, per poi tornare a crescere con qualche oscillazione nelle ultime epoche di training.

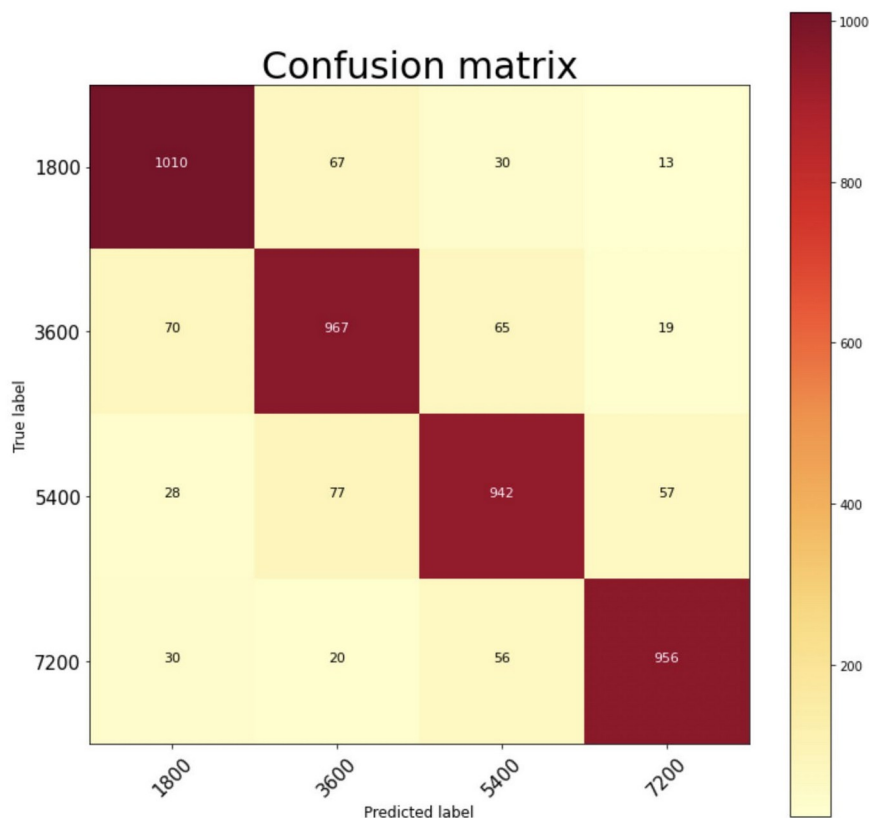


Figura 4.8 Matrice di confusione del modello

La matrice di confusione del modello restituisce una rappresentazione dell'accuratezza dei risultati e permette di analizzare gli errori con l'obiettivo di migliorare il modello (Figura 4.8).

Sull'asse delle ascisse troviamo le "predicted labels" mentre le "true labels" sono situate sull'asse delle ordinate.

Ciascuna i -esima cella rappresenta il numero di casi in cui la CNN ha classificato la predicted label come una true label.

Nella diagonale troviamo le corrette classificazioni che, già a primo impatto, rappresentano la maggioranza.

Inoltre, è possibile ottenere due valori di accuratezza: Producer Accuracy ovvero il numero di valori di una classe correttamente classificati / numero di valori appartenenti alla classe in questione e User Accuracy ovvero il numero di valori di una classe correttamente classificati / numero di valori classificati come appartenenti alla classe.

La Producer Accuracy esprime quindi la probabilità che una certa immagine appartenente a una classe sia classificata come tale. Quindi, considerando la classe 1800, su tutte le volte che l'immagine apparteneva alla classe in questione, quante volte è stata classificata come tale? La Producer Accuracy è complementare all'errore di omissione, il quale dà una stima dei pattern omessi dalla corretta classe.

La User Accuracy invece, è una stima della percentuale di elementi correttamente attribuiti ad una data classe rispetto al totale degli elementi attribuiti a quella classe. Quindi, considerando sempre la classe 1800, su tutte le volte che il modello ha classificato l'immagine come appartenente a tale classe, quante volte effettivamente l'immagine apparteneva a questa classe? La User Accuracy è complementare all'errore di commissione che dà una misura dei pattern incorrettamente inclusi in una categoria.

Nella tabella 4.3 sono riportati i valori di Producer e User Accuracy per le 4 categorie.

CLASSE	PRODUCER ACCURACY	USER ACCURACY
1800	90.18%	88.75%
3600	86.26%	85.5%
5400	85.33%	86.18%
7200	90%	91.48%

Tabella 4.3 Producer e User Accuracy

5. DISCUSSIONE

Lo scopo della presente ricerca è stato quello di prevedere un attacco epilettico, attraverso un modello di Deep Learning, sulla base del tracciato elettroencefalografico di un paziente.

La disponibilità di grandi set di dati EEG ha ampliato la portata dell'utilizzo di architetture di Deep Learning per la creazione di modelli robusti per la classificazione, in quanto, queste reti neurali possono risultare molto vantaggiose con l'utilizzo di Big Data (Cherian & Kanaga, 2022).

Durante la progettazione di un classificatore per la previsione di una seizure è possibile utilizzare un approccio patient-independent o patient-dependent (Cherian & Kanaga, 2022).

Nel primo caso, gli studi mirano a progettare un classificatore in grado di riconoscere le crisi in più pazienti; la rete neurale, quindi, utilizza l'intero set di dati per progettare una funzione predittiva in grado di identificare l'attacco epilettico in più soggetti (Dissanayake et al., 2020).

Nel secondo caso invece, a causa dell'elevata variabilità dei dati EEG, si genera un classificatore per ogni paziente. Si crea quindi un unico framework di classificazione, personalizzato per ogni individuo.

Nel presente studio abbiamo adottato un approccio patient-independent.

Sebbene infatti i disegni sperimentali patient-dependent possano garantire prestazioni migliori rispetto ai modelli indipendenti dal paziente (poiché, a differenza di quest'ultimi, non devono adattarsi all'elevata variabilità inter-soggetto), tuttavia richiedono una grande mole di dati preictali per addestrare adeguatamente i modelli neurali.

È bene evidenziare che spesso i dataset clinici comprendono un numero variabile di registrazioni per ciascun paziente e, pertanto, essere in grado di apprendere una funzione globale che consideri l'intero set di dati è importante per garantire una maggiore generalizzazione.

Il dataset fornito dall'ospedale di Conegliano contiene le registrazioni EEG di 20 pazienti. Per ciascuno di essi è stato registrato un numero di canali differente e campionati ad una

sampling rate diversa (256 Hz, 512 Hz, 1024 Hz). Infatti, il segnale EEG di 11 pazienti è stato campionato a 256 Hz, quello di 6 soggetti a 512 Hz, mentre i restanti 3 a 1024 Hz. Sebbene ad una prima analisi questi fattori possano rappresentare un bias per la classificazione degli spettrogrammi, il potenziale errore che il dataset porta con sé è stato ridotto utilizzando alcuni accorgimenti sperimentali al fine di limitare eventuali distorsioni.

Per evitare registrazioni EEG troppo pesanti abbiamo utilizzato un downsampling comune a 256 Hz in modo tale da creare un dataset uniforme, costituito da tracciati elettroencefalografici dotati dello stesso numero di sampling point per intervallo di tempo.

D'altro canto, siccome le registrazioni di ciascun paziente contengono un numero di canali differente, sono stati scelti canali comuni tra i soggetti per avere tracciati EEG bilanciati.

Questi accorgimenti che sono stati adottati dovrebbero essere utili per ridurre il bias potenziale che il dataset porta con sé; tuttavia, in una futura estensione della ricerca, sarebbe opportuno confrontare i risultati ottenuti dalle presenti architetture neurali con quelli ottenibili dai medesimi modelli applicati a dataset con un numero fisso di canali e campionati alla stessa sampling rate (ad esempio CHB-MIT).

Nella ricerca proposta da Troung et al. (2017) è stato utilizzato il Database iEEG dell'ospedale di Friburgo ed il dataset EEG CHBMIT per allenare e testare la CNN. In analogia con il nostro disegno sperimentale, hanno utilizzato un sampling rate pari a 256 Hz e, attraverso la STFT, hanno trasformato i dati EEG grezzi in una matrice bidimensionale.

Diversamente da quanto proposto nel nostro esperimento, questa immagine viene passata attraverso la CNN per l'apprendimento delle caratteristiche e la classificazione degli stati preictali ed interictali. Come evidenziato in Figura 5.1 (a) per la valutazione delle prestazioni del modello hanno impostato l'indice "seizure prediction horizon" (SPH) a 5 minuti e il "seizure occurrence period" (SOP) a 30 minuti ed hanno utilizzato la sensibilità e il false prediction rate come metriche di valutazione (Rasheed et al., 2020).

Nel nostro disegno sperimentale, d'altro canto, non abbiamo tenuto in considerazione questi due indici, ma abbiamo voluto generare un modello che andasse a classificare le quattro categorie di appartenenza del segnale antecedente alla crisi.

Inoltre, nel momento in cui la durata del periodo preictale (così come quella del periodo post-ictale) varia da paziente in paziente, abbiamo deciso di delineare quattro categorie che non tengano conto della differenza tra il segnale preictale ed interictale, in modo tale da costruire un design sperimentale che possa essere libero da questo vincolo e, quindi, maggiormente adattabile ad un approccio patient-independent.

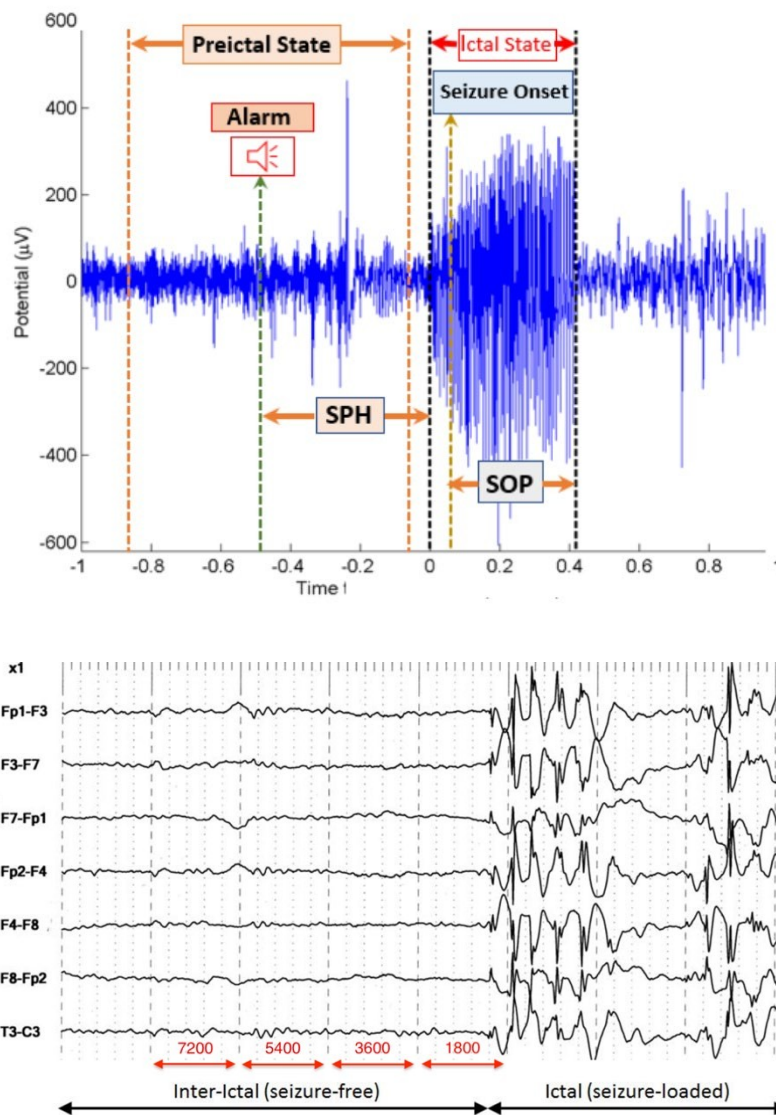


Figura 5.5 Disegni sperimentali a confronto. Figura a) Concetto di SOP e SPH. Figura b) Divisione dataset in 4 categorie precedenti alla seizure

L'utilizzo della Short Time Fourier Transform (STFT) è stata necessaria per trasformare i dati grezzi in una matrice bidimensionale, vale a dire una finestra EEG avente due dimensioni: frequenza e tempo.

Creare un'immagine a partire dal dato sotto forma di onde cerebrali è utile per permettere ad un filtro di convoluzione bidimensionale di scorrere in tutta la STFT per raccogliere i cambiamenti sia nella frequenza che nel tempo dei segnali EEG.

Siccome la registrazione del segnale è avvenuta in un ambiente aperto, ovvero in una camera in cui i pazienti sono liberi di muoversi, è stato opportunamente applicato un notch filter a 50 e 100 Hz per limitare il rumore di rete che potrebbe ripercuotersi nelle sue armoniche, essendo il segnale EEG registrato in ambiente non schermato.

Nei pazienti con epilessia si possono inoltre osservare oscillazioni aperiodiche ad elevata frequenza (250 – 600 Hz). Questo fenomeno, che prende il nome di fast ripples (FR), si verifica soprattutto nelle prime fasi dell'epilettogenesi.

Sebbene le dinamiche alla base della generazione dei fast ripples non siano ancora ben comprese, sembra che queste onde possano essere un biomarcatore del focus di generazione della crisi.

Perché, quindi, filtrare con un passa-basso a 125 Hz?

Il segnale elettroencefalografico di un paziente con epilessia è molto eterogeneo in termini di frequenza e caratteristiche: possono verificarsi i fast ripples, poli-punte bilaterali e anomalie delle onde a una frequenza da 4 a 6 Hz, scariche di punte e onde lente bilaterali a una frequenza di 3/secondo e così via.

Le registrazioni della maggior parte dei pazienti sono campionate a 256 Hz, quindi, applicare un downsampling a 256 Hz per tutti i soggetti componenti il dataset esclude a priori i fast ripples dall'analisi.

Inoltre, molti dataset pubblicati per il "seizure prediction task" hanno una frequenza di campionamento pari o inferiore a 256 Hz (CHB-MIT, MSSM, Freiburg, Bonn).

Utilizzare un filtro low-pass a 125 Hz è quindi possibile perché non solo rispetta il criterio di Nyquist, ma permette anche di includere le principali onde cerebrali (tabella 5.1).

Frequency band	Frequency (Hz)	Brain State
Delta	1–4	Deepest meditation, dreamless sleep
Theta	4–8	Deeply relaxed, meditation
Alpha	8–13	Very relaxed, passive attention, calm
Beta	13–30	Alert, attentive
Gamma	30–80	Reduced consciousness, Stimulus

Tabella 5.3 Onde cerebrali (Tsipouras, 2019)

Come anticipato nei capitoli precedenti, per prevedere la crisi epilettica è stato utilizzato un modello Deep Learning.

Le architetture di DL rappresentano il risultato di progressi nella ricerca del machine learning e sono state utilizzate in molti domini medici, quali l'imaging clinico, la biologia computazionale e la previsione delle malattie (Rasheed et al., 2020).

Infatti, per utilizzare un approccio di ML è necessaria una fase di “feature extraction”. Questo processo è dispendioso in termini temporali a causa della presenza di rumore e artefatti nei dati e, sebbene queste architetture possano fornire una buona accuratezza, la possibilità di creare un modello generalizzabile non è molto elevata. Lo step di “feature extraction” non è potente tanto quanto i modelli di deep learning per la previsione delle crisi epilettiche perché i dati EEG non solo variano tra i pazienti, ma variano anche per lo stesso paziente nel tempo.

D'altra parte, gli algoritmi DL apprendono automaticamente le feature che risultano essere più distintive e robuste (Andreas et al., 2016).

Sebbene, al momento, non sia ancora disponibile una previsione che fornisca un'accuratezza pari al 100%, con le attuali prestazioni che si stanno via via raggiungendo con i modelli di Machine Learning e soprattutto Deep Learning, sembra possibile fornire ai pazienti un preavviso in modo tale che possano prendere alcune precauzioni al fine di evitare o diminuire l'impatto di un attacco epilettico.

Nel seguente studio sono stati addestrati due modelli di IA: VGG16 e CNN

Per entrambe le architetture, prima dell'addestramento finale è stata svolta un'operazione di hyperparameters tuning con lo scopo di trovare la combinazione di iperparametri che producesse una migliore accuratezza e un minor errore.

Per quanto concerne la VGG, come precedentemente accennato, i risultati non sono soddisfacenti: la funzione d'errore decresce troppo lentamente e l'accuratezza non mostra considerevoli miglioramenti epoca dopo epoca.

Inoltre, nonostante un training più robusto su 300 epoche e in assenza del weight decay (parametro utile soprattutto per evitare l'overfitting) viene raggiunta un'accuratezza massima del 30.61% all'epoca 291.

Questi risultati si riflettono poi sul test set (e sulla relativa matrice di confusione) che mette in luce le performance di una rete che classifica in modo casuale gli spettrogrammi entro le 4 classi producendo, infatti, un'accuratezza pari al 25.12%.

Questi esiti possono essere dovuti ad una combinazione di iperparametri che non producono buoni risultati. Quindi, in una futura estensione della ricerca sarebbe opportuna un'esplorazione più sistematica degli iperparametri e un debug del codice per analizzare eventuali errori.

D'altro canto, per quanto riguarda la CNN i risultati sono soddisfacenti.

Analizzando infatti la tabella 4.1 del capitolo precedente, emerge subito come il learning rate sia l'iperparametro che maggiormente influisce nel determinare la performance del modello. Diversamente, sebbene batch size e GAP abbiano sicuramente un impatto sull'accuratezza, non è presente un trend crescente o decrescente che permetta di stabilire una variazione di performance all'aumentare o diminuire del batch size o nell'includere o meno il GAP.

Considerando, ad esempio, l'accuratezza prodotta al variare del learning rate, emerge un trend decrescente per quest'ultima all'aumentare del valore del tasso di apprendimento. Questo effetto, visibile sia al variare del batch size sia in presenza o assenza del GAP, è maggiormente pronunciato per un learning rate pari a 5×10^{-3} e 10^{-2} .

Come accennato precedentemente il batch size ha un effetto sull'accuratezza, ma non si verifica un andamento crescente o decrescente al variare di questo iperparametro. Prendendo in esame, ad esempio, le prime tre epoche della tabella 4.1, performance migliori si ottengono per un basso valore di batch size mentre, variando il learning rate si inverte la tendenza in quanto prestazioni migliori sono conseguite da valori di batch size più elevati.

Infine, sebbene la miglior combinazione di iperparametri includa anche la presenza del Global Average Pooling, la sua presenza o meno non determina un sostanziale aumento dell'accuratezza al variare degli altri parametri.

Inoltre, esaminando i grafici di loss e accuracy emerge un leggero overfitting nelle epoche finali. L'Overfitting è quel fenomeno per cui la loss nel training diminuisce, mentre nel validation set inizia ad aumentare. Questo processo si verifica in quanto la rete apprende troppo dettagliatamente le caratteristiche dei pattern di addestramento, a discapito della capacità di generalizzazione del modello. Sarebbe quindi opportuno utilizzare una procedura di early stopping in quanto, dal grafico dell'accuratezza, emerge che quest'ultima non sembra aumentare in modo sostanziale dopo l'epoca 60 e, dal grafico della loss, si evince che la funzione d'errore inizi ad aumentare dopo tale epoca.

In aggiunta, i risultati che emergono dalla matrice di confusione dopo aver testato la rete sul test set sono soddisfacenti: gli errori, seppur pochi rispetto all'ammontare del numero di pattern che la rete classifica, si verificano principalmente per le categorie adiacenti a quella target. Ad esempio, prendendo in considerazione la classe 3600, la rete, su un totale di 1131 pattern appartenenti alla categoria, 77 volte classifica lo spettrogramma nella classe 1800 e 67 volte in 5400. Un errore molto inferiore (20 pattern) si verifica nella categoria 7200, in quanto non adiacente alla classe 3600.

Considerando User e Producer Accuracy, queste mostrano valori molto simili tra loro. Ciò significa che il classificatore raggiunge buone prestazioni. Ad esempio, esaminando la classe 1800, la Producer Accuracy è 90.18% mentre la User Accuracy è 88.75%; questo significa che quando l'immagine appartiene alla categoria 1800, nel 90.18% dei casi la CNN classifica correttamente l'immagine e nell'88.75% delle volte che ha classificato un'immagine come appartenente alla classe 1800, lo spettrogramma appartiene effettivamente a questa categoria. Questo risultato dimostra che un buon classificatore produce valori simili tra queste due accuratèzze.

In questo progetto, quindi, abbiamo proposto un nuovo approccio che utilizzi la CNN con funzionalità ingegneristiche minime.

Sebbene il dataset abbia mostrato qualche criticità, la creazione di un modello di classificazione che categorizzi le immagini pre-ictali entro quattro classi ha fornito

risultati robusti. Fondamentali sono state le operazioni di preprocessing volte a ridurre i potenziali bias nei dati.

Tale metodo, quindi, ha mostrato una buona generalizzazione offrendo, in futuro, ai pazienti l'opportunità di possedere un dispositivo di previsione delle crisi che possa aiutarli ad avere una vita più gestibile (Truong et al., 2018).

Nel caso di allarme generato dal dispositivo in questione è bene preparare il paziente ad un possibile attacco. È fondamentale, quindi, allontanare oggetti spigolosi come tavoli e sedie, utilizzare un supporto morbido sotto al capo, allentare la cravatta o togliere gli indumenti stretti, evitare di mettersi alla guida o compiere azioni che possano provocare danni a sé o agli altri.

Sembrerebbe, quindi, che in un futuro forse non troppo lontano la possibilità avere un dispositivo che preveda la crisi epilettica sulla base del segnale EEG possa essere, oltre che utile, uno strumento salva vita.

BIBLIOGRAFIA

- [1] Abdi-Sargezeh B., Valentin A., Alarcon G. and Sanei S., (2021), " Incorporating Uncertainty in Data Labeling Into Detection of Brain Interictal Epileptiform Discharges From EEG Using Weighted optimization," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 1000-1004
- [2] Abrol, A., Fu, Z., Salman, M. et al. (2021) “ Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning “. Nat Commun 12, 353
- [3] Acharya U. R., Oh S. L., Hagiwara Y., Tan J. H., Adeli H., (2017), “ Deep convolutional neural network for automated detection and diagnosis of seizure using EEG signals ”. Computers in Biology and Medicine
- [4] Allen Hauser W., John F Annegers, and Walter A Rocca. (1996), “Descriptive epidemiology of epilepsy: contributions of population-based studies from rochester, Minnesota”. In Mayo Clinic Proceedings, volume 71, pp. 576–586.
- [5] Andreas A., Loukianos S., Clive C. T., and Saeid S. (2016), “Deep learning for epileptic intracranial eeg data”. In 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (*MLSP*), pp 1–6. IEEE, 2016.
- [6] Bear M.F., Barry W. Connors B. W., Paradiso M., (2016), “Neuroscienze esplorando il cervello”
- [7] Cherian R., Kanaga E.G. (2022), “ Theoretical and methodological analysis of EEG based seizure detection and prediction: An exhaustive review ”. Journal of Neuroscience Methods
- [8] Commission on Classification and Terminology of the International League Against Epilepsy. Proposal for revised classification of epilepsies and epileptic syndromes. *Epilepsia*, 30(4):389–399, 1989.
- [9] Craik A., He Y., Contreras-Vidal J. (2019), “ Deep learning for electroencephalogram (EEG) classification tasks: a review ”. Journal of Neural Engineering, 16 (3)

- [10] Dissanayake T., et al. (2021) “Deep Learning for Patient-Independent Epileptic Seizure Prediction Using Scalp EEG Signals” . IEEE sensors journal, vol. 21, no. 7. Pp 9367 – 938
- [11] DK Angeles et al. (1981) ”Proposal for revised clinical and electroencephalographic classification of epileptic seizures”. *Epilepsia*, 22(4): 489–501.
- [12] Eberlein M., Hildebrand R., Tetzlaff R., Hoffmann N., Kuhlmann L., B., Müller J. (2018), “ Convolutional Neural Networks for Epileptic Seizure Prediction “. IEEE International Conference on Bioinformatics and Biomedicine (BIBM)
- [13] Fukushima K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4), 193-202.
- [14] Galanopoulou A., Buckmaster P., Staley K., (2012), “Identification of new epilepsy treatments: Issues in preclinical methodology”. Official journal of the international league against epilepsy
- [15] Goodfellow I., Bengio Y., Courville A. (2016), “Deep learning book”. MIT Press, 2016
- [16] Grace W. Lindsay, (2020) “Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future”
- [17] Hubel, D. H., & Wiesel, T. N. (1962). “Receptive fields, binocular interaction and functional architecture in the cat's visual cortex”. *The Journal of physiology*, 160(1), 106-154
- [18] Jiang X., Bian G., Tian Z., (2019), “Removal of Artifacts from EEG Signals: A Review”. Institute of Automation, Chinese Academy of Science, Beijing
- [19] Ko. DY, Benbadis. RS, (2016), “Epilepsia e convulsioni”. Medscape
- [20] LeCunn et al. (1989) “Handwritten Digit Recognition With A Back-Propagation Network”, NeurIPS conference
- [21] Lundervold, A. S., & Lundervold, A. (2019). “ An overview of deep learning in medical imaging focusing on MRI “. *Zeitschrift für Medizinische Physik*, 29(2), 102-127.
- [22] Maiwald T., Winterhalder M., Aschenbrenner-Scheibe R., Voss H. U., Schulze-Bonhage A., Timmer J. (2004), “ Comparison of three nonlinear seizure prediction

- methods by means of the seizure prediction characteristic “. *Physica D: Nonlinear Phenomena*, 194, pp. 357-368
- [23] Ohiraa J., Yoshimuraa H., Morimoto T., Ariyoshid K., Kohara N., (2018), “ Factors associated with the duration of the postictal state after a generalized convulsion “. *Seizure: European Journal of Epilepsy* 65, pp. 101-105
- [24] OShea, A., Ahmed, R., Lightbody, G., Mathieson, S., Pavlidis, E., Lloyd, R., ... & Temko, A. (2021). “Deep learning for EEG seizure detection in preterm infants ”. *International Journal of Neural Systems*
- [25] Ozdemir, M. A., Cura, O. K., & Akan, A. (2021). “ Epileptic eeg classification by using time-frequency images for deep learning “. *International Journal of Neural Systems*, 2150026.
- [26] Rasheed K., Qayyum A., Qadir J., (2020), “ Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review ”. *IEEE Reviews in Biomedical Engineering* (14)
- [27] R. Battistin (2022), “*Neurologia Clinica*”. Società Editrice Esculapio
- [28] Savarino V., Craxi A., Prisco D., et al. (2012), “*Manuale medico di diagnostica e terapia roversi*”.
- [29] Shoeibi A., Khodatars M., Ghassemi N. (2021), “ Epileptic Seizures Detection Using Deep Learning Techniques: A Review “. *International Journal of Environmental Research and Public Health* 18(11):5780
- [30] Shoka A., Dessouky M., El-Sherbeny A. S., El-Sayed A. (2019), “Literature Review on EEG Preprocessing, Feature Extraction, and Classifications Techniques “. *Menoufia Journal of Electronic Engineering Research*
- [31] Smith S., (2005), “ Eeg in the diagnosis, classification, and management of patients with epilepsy “. *J Neurol Neurosurg Psychiatry* (76)
- [32] Thangavel, P., Thomas, J., Peh, W. Y., Jing, J., Yuvaraj, R., Cash, S. S., ... & Dauwels, J. (2021). “ Time-Frequency Decomposition of Scalp Electroencephalograms Improves Deep Learning-Based Epilepsy Diagnosis “. *International Journal of Neural Systems*, 31(08), 2150032.
- [33] Truong N. D., Nguyen A. D., Kuhlmann L., Bonyadi M. R., Yang J., Ippolito S., Kavehei O., (2018), “Convolutional neural networks for seizure prediction using

intracranial and scalp electroencephalogram “. Neural Networks, 105, pp. 104-111

- [34] Tsiouris K. M., Pezoulas V. C., Zervakis M., Konitsiotis S., Koutsourisa D. D., Fotiadis D. (2018), “ A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals “. Computers in Biology and Medicine (99)
- [35] <https://machinelearningmastery.com/how-to-use-transfer-learning-when-developing-convolutional-neural-network-models/>
- [36] <https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c>