

Università Politecnica delle Marche

Facoltà di Ingegneria

Dipartimento di Ingegneria dell'Informazione

Corso di Laurea Magistrale in Ingegneria Elettronica



Tesi di Laurea

Identificazione di apnee notturne del sonno tramite
segnali audio e algoritmi avanzati di deep learning

Identification of sleep apnea using audio signals and
advanced deep learning algorithms

Relatore

Prof. Stefano Squartini

Correlatori

Dott. Carlo Aironi

Dott.ssa Lucia Migliorelli

Candidato

Davide Lillini

Sessione di Laurea Ottobre 2023
Anno Accademico 2022/2023

Indice

1	Introduzione	5
1.1	Apnee ostruttive nel sonno	5
1.2	Apnee centrali del sonno	10
2	Stato dell'arte e scopo della tesi	12
2.1	Metodologie diagnostiche validate clinicamente	12
2.2	Scopo della tesi	15
2.3	Identificazione automatica di apnee notturne da audioregistrazioni . . .	16
3	Materiali e metodi	20
3.1	Architettura di rete neurale impiegata	20
3.2	Dataset	23
3.3	Pre-processing del dataset	28
3.4	Creazione del dataset di addestramento e impostazioni di addestramento	31
3.5	Metriche valutative utilizzate	33
3.5.1	Accuratezza	34
3.5.2	Recall	34
3.5.3	Precision	35
3.5.4	Precision-Recall Curve	35
3.6	Validazione del modello	36
3.6.1	Apnee ostruttive	37
3.6.2	Apnee Centrali	39
3.6.3	Preparazione del test set	41
4	Risultati e discussione	43
4.1	Apnee ostruttive	43
4.2	Apnee Centrali	47
5	Conclusioni e sviluppi futuri	50

Elenco delle figure

7figure.caption.5

1.2	Esempio di disposizione degli elettrodi per EGG (1).	9
2.1	Disposizione della sensoristica durante un analisi polisonnografica.	14
2.2	Esempio di schermata video relativa ad un'analisi polisonnografica in presenza di un'apnea ostruttiva (2).	15
2.3	Diagramma di flusso del metodo proposto	16
3.1	Diagramma a blocchi CRNN.	21
3.2	Setup della postazione usata per l'acquisizione dei dati (3).	25
3.3	Microfono a contatto Clockaudio CTH100.	26
3.4	Esempio di utilizzo del microfono a contatto Clockaudio CTH100.	26
3.5	Distribuzione dei tipi di apnea all'interno del dataset (3).	28
3.6	Durata in s degli eventi annotati nel dataset e relativa numerosità (3).	28
3.7	Grafico densità di probabilità-durata degli eventi annotati.	29
3.8	In rosso la maschera generata per identificare le apnee.	30
3.9	Esempio di curva precision-recall.	36
3.10	Accuratezza in fase di validazione	37
3.11	Loss in fase di validazione	38
3.12	Recall in fase di validazione	38
3.13	Accuratezza in fase di validazione	39
3.14	Loss in fase di validazione	40
3.15	Loss in fase di validazione	40
3.16	Filtraggio falsi positivi	41
4.1	Curva precision-recall	44
4.2	Curva precision-recall dopo il post-processing	45
4.3	Esempio di spettrogramma con presenza di rumore invalidante	46
4.4	Esempio di spettrogramma con una presenza di rumore non invalidante	46
4.5	Curva precision-recall per apnee centrali	48
4.6	Curva precision-recall per apnee centrali dopo il post-processing	49

Sommario

La sindrome delle apnee notturne rappresenta un complesso di disturbi respiratori che si manifestano durante il sonno, caratterizzati da episodi di ostruzione totale o parziale delle vie aeree superiori. Questi episodi portano a una temporanea interruzione/diminuzione del flusso d'aria e a una conseguente riduzione dell'ossigenazione. La gravità di tali disturbi può variare notevolmente, spaziando da forme lievi a manifestazioni gravi, con potenziali impatti negativi sulla salute di chi ne è affetto. La corretta identificazione e la diagnosi delle apnee notturne sono cruciali per prevenire possibili conseguenze a lungo termine, tra cui l'insorgenza di ipertensione, malattie cardiovascolari e disfunzioni cognitive.

Dal punto di vista demografico, le apnee notturne costituiscono un problema di portata globale. Le stime indicano che l'incidenza della sindrome delle apnee ostruttive del sonno, la forma più comune di apnee notturne, si attesta tra il 9% e il 24% negli adulti. Un aspetto interessante riguarda la differenza di rischio tra i sessi, con gli uomini che presentano una prevalenza maggiore rispetto alle donne in un rapporto di circa 2:1. Inoltre, vi è un chiaro incremento dell'incidenza delle apnee notturne con l'aumentare dell'età, raggiungendo il picco tra i 40 e i 60 anni. È importante sottolineare che, sebbene l'associazione tra l'età avanzata e le apnee notturne sia ben documentata, queste ultime non sono limitate solamente agli anziani, ma possono verificarsi anche in soggetti giovani e persino nei bambini.

Per fronteggiare la sfida rappresentata dalla diagnosi delle apnee notturne, è stato avviato uno sforzo significativo nello sviluppo di sistemi di identificazione automatica. Questi sistemi mirano a semplificare la procedura di polisonnografia, tradizionalmente utilizzata per diagnosticare le apnee notturne. La polisonnografia richiede l'analisi approfondita del sonno attraverso sensori elettrofisiologici durante una notte di riposo, ma è un processo costoso, invasivo e che richiede la permanenza del paziente in ospedale. Ciò rappresenta un onere per il sistema sanitario e comporta tempi di attesa prolungati per la diagnosi e il trattamento.

L'integrazione di tecnologie all'avanguardia, come l'apprendimento automatico e l'analisi dei segnali biologici, ha permesso lo sviluppo di algoritmi avanzati in grado di identificare gli episodi di apnea durante il sonno. Questi sistemi possono essere

implementati attraverso dispositivi indossabili o applicazioni mobili, consentendo un monitoraggio non invasivo e continuo delle apnee notturne anche in contesti domestici. Questo approccio potrebbe rivoluzionare il processo diagnostico, rendendolo più accessibile, meno oneroso e consentendo una diagnosi precoce.

Alla luce delle informazioni fornite, lo scopo principale di questa tesi è sviluppare un metodo basato sul deep learning per l'identificazione automatica degli episodi di apnea di tipo centrale ed ostruttivo. Questo metodo si baserà unicamente sui dati registrati durante i cicli di sonno dei pazienti, ottenuti attraverso l'uso di microfoni tracheali.

Le apnee notturne rappresentano un problema di salute diffuso con importanti implicazioni per la salute generale. La distribuzione demografica delle apnee notturne evidenzia l'importanza di considerare fattori come età e genere nella strategia di identificazione e trattamento. L'evoluzione verso sistemi di identificazione automatica offre una prospettiva promettente per ottimizzare la procedura di diagnosi, ridurre i costi e migliorare l'accessibilità alla cura. Tuttavia, ulteriori ricerche sono necessarie per raffinare questi sistemi, garantirne l'accuratezza e valutarne l'efficacia a lungo termine.

Capitolo 1

Introduzione

In questo capitolo verrà approfondita la sindrome delle Apnee Notturme del Sonno dal punto di vista clinico.

Il termine "disturbi respiratori del sonno" di cui fanno parte anche le apnee del sonno comprende un complesso di condizioni caratterizzate da alterazioni nella respirazione durante il periodo di sonno. In molte circostanze, questa condizione è associata a una restrizione o occlusione delle vie aeree superiori, in particolare la faringe. I disturbi respiratori del sonno possono variare notevolmente, da una ostruzione parziale e intermittente delle vie aeree senza conseguenze sul sonno, noto come roncopatia, fino a una forma più grave caratterizzata da apnee ricorrenti, accompagnate da episodi frequenti di bassa saturazione di ossigeno nel sangue e risvegli, che provocano una significativa frammentazione del sonno e sonnolenza durante il giorno. Nel contesto dei disturbi respiratori del sonno, rientrano diversi tipi di disturbi, tra cui le apnee Centrali del sonno (CSA), caratterizzate da interruzioni periodiche del sonno senza ostruzione delle vie aeree, spesso riscontrate negli adulti affetti da scompenso cardiaco e le apnee ostruttive del sonno anche esse caratterizzate interruzioni periodiche del sonno dovute all'ostruzione totale o parziale delle vie aeree.

1.1 Apnee ostruttive nel sonno

L'Apnea Ostruttiva del Sonno (SAS Ostruttiva) è una condizione diffusa, ma spesso sotto-diagnosticata. Nei paesi sviluppati, si stima che colpisca tra il 3% e il 7% degli uomini di mezza età e il 2% al 5% delle donne. Questa condizione è caratterizzata da frequenti ostruzioni delle vie aeree superiori durante il sonno, causando ripetute interruzioni della respirazione, riduzione dell'ossigeno nel sangue e risvegli notturni. Questa interruzione del sonno porta a sonnolenza diurna e, a lungo termine, può

causare problemi cognitivi e aumentare il rischio di malattie cardiovascolari¹

La manifestazione clinica e i criteri diagnostici dei disturbi respiratori del sonno differiscono tra adulti e bambini. La SAS Ostruttiva è più comune in alcuni gruppi di persone, come gli individui obesi e quelli con condizioni mediche specifiche, come la sindrome di Down.

L'Apnea Ostruttiva del Sonno è caratterizzata da episodi di ostruzione delle vie aeree superiori, che possono essere classificati come "apnee" quando le vie aeree sono completamente occluse e "ipopnee" quando l'occlusione è parziale. Un'apnea ostruttiva è considerata come un'interruzione del flusso d'aria che persiste per almeno 10 secondi, nonostante siano presenti tentativi continui di respirare. Spesso, al termine di queste apnee/ipopnee, si verificano cambiamenti nel segnale elettroencefalografico (EEG), indicando risvegli, e una diminuzione dei livelli di ossigeno nel sangue. Tuttavia, nella maggior parte dei casi, questi risvegli sono brevi e il paziente non è consapevole di essi (4).

La definizione esatta di ipopnea può variare a seconda del tipo di strumentazione utilizzata per monitorare la respirazione, ma in linea di principio, secondo l'American Academy of Sleep Medicine (AASM), coinvolge una riduzione del 30% al 50% nei movimenti toraco-addominali rispetto ai livelli basali stabili per almeno 10 secondi. Le linee guida attuali dell'AASM del 2012 richiedono anche la presenza di una riduzione del 3% nei livelli di saturazione dell'ossigeno o un risveglio. Tuttavia, in alcuni centri medici, possono essere in uso definizioni di ipopnea diverse.

La faringe rappresenta il luogo principale dell'occlusione delle vie aeree superiori durante un evento di apnea ostruttiva del sonno. In generale, qualsiasi alterazione patologica o variazione anatomica normale che restringa le vie aeree superiori durante lo stato di veglia può predisporre a episodi di apnee o ipopnee ostruttive durante il sonno (6).

Durante l'atto di inspirazione, la pressione dell'aria nella faringe è inferiore a quella atmosferica, e le dimensioni del lume faringeo sono influenzate dall'equilibrio tra le forze di restringimento, che derivano dalla pressione negativa, e le forze dilatanti, generate dai piccoli muscoli presenti nelle vie aeree superiori. Questi muscoli si contraggono durante ogni inspirazione e di solito mantengono stabile la parete molle della faringe. All'inizio del sonno, si verifica una riduzione dell'ampiezza del lume faringeo e una diminuzione dell'attività dei muscoli delle vie aeree superiori, entrambi amplificati nei casi di apnee ostruttive. Inoltre, le condizioni delle mucose di superficie possono influire sull'apertura delle vie aeree, specialmente nei soggetti con mucose infiammate a causa di traumi ripetuti e della conseguente perdita di sensibilità.

¹Url: <https://www.area-c54.it/public/disturbi%20respiratori%20del%20sonno.pdf>

Ogni episodio di apnea o ipopnea termina con un risveglio, accompagnato da un brusco aumento della frequenza cardiaca e della pressione arteriosa. In molti individui, questo aumento della pressione arteriosa persiste per giorni, aumentando il rischio di sviluppare malattie cardiovascolari e ictus.

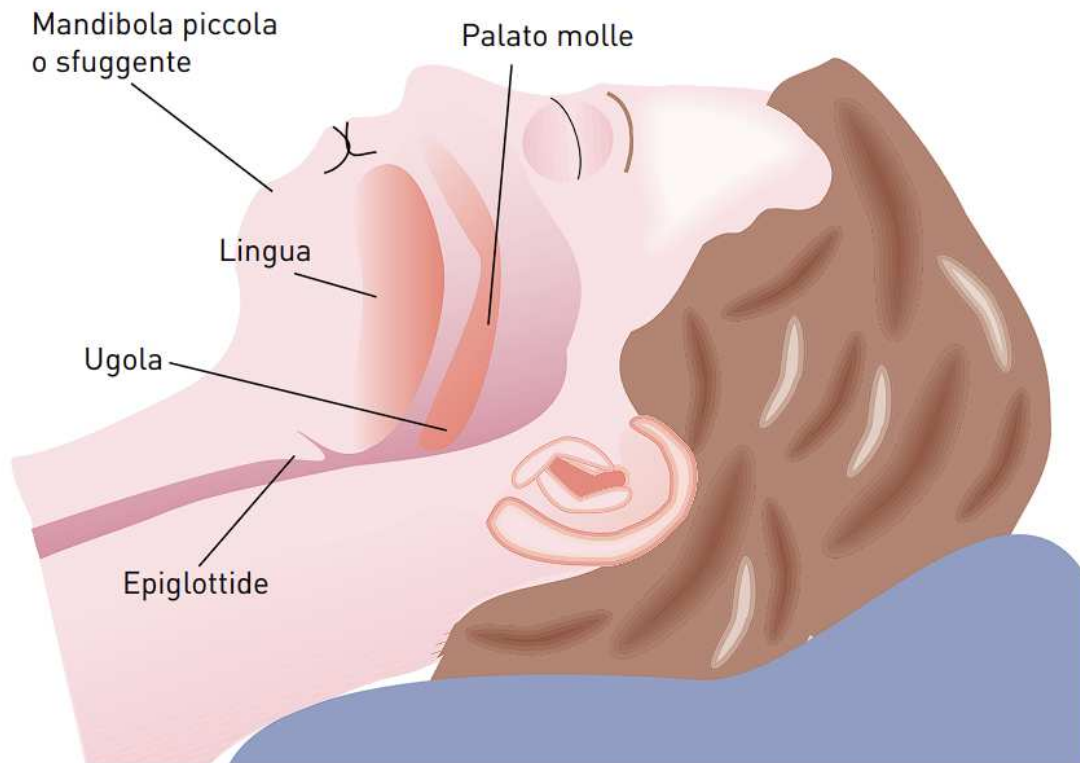


Figura 1.1: Sezione della faringe posteriore, che mostra il meccanismo di ostruzione delle vie aeree superiori durante il sonno².

La prevalenza della SAS Ostruttiva aumenta con l'età e raggiunge un livello stabile dopo i 60 anni. Tuttavia, dati recenti basati su uno studio trasversale con oltre 5.000 partecipanti hanno rivelato che una percentuale significativa di persone di età superiore ai 70 anni continua a manifestare i sintomi della malattia (7).

E' da precisare inoltre che l'obesità è il fattore predisponente più comune, ma i pazienti possono anche presentare altri fattori che contribuiscono al restringimento delle vie aeree, come una lingua voluminosa, tonsille ingrossate, un'eccessiva flaccidità del tessuto faringeo o una mandibola retrusa (figura 1.1).

Numerose ricerche infatti hanno confermato l'associazione tra obesità e SAS Ostruttiva, con il 60-90% dei pazienti affetti da SAS Ostruttiva che presentano un'elevata obesità, definita da un indice di massa corporea (BMI) superiore a 30 kg/m². Inoltre, l'obesità di tipo centrale, caratterizzata da un elevato rapporto vita-fianchi o da una circonferenza del collo ampia, sembra correlare meglio con l'SAS Ostruttiva rispetto al BMI, anche in individui con un BMI considerato normale (10).

La SAS Ostruttiva è più comune negli uomini rispetto alle donne, il che può essere attribuito alle differenze anatomiche e funzionali delle vie aeree superiori, alle caratteristiche cranio-facciali, all'accumulo di grasso e alle risposte ventilatorie al risveglio durante il sonno. La prevalenza della SAS Ostruttiva aumenta nelle donne in post-menopausa e diminuisce con la terapia ormonale, ma tende ad aumentare durante la gravidanza, soprattutto nel terzo trimestre.

I parenti di primo grado dei pazienti con SAS Ostruttiva hanno un rischio maggiore di sviluppare la malattia, e sono rilevanti anche i fattori genetici legati a caratteristiche cranio-facciali, obesità e distribuzione del grasso corporeo. Condizioni congenite che influenzano lo sviluppo cranio-facciale, come la sindrome di Marfan, la sindrome di Down e la sequenza di Pierre Robin, aumentano il rischio di SAS Ostruttiva, così come acromegalia e ipotiroidismo.

Il fumo è associato a una maggiore prevalenza di russamento e SAS Ostruttiva, mentre l'assunzione di alcol può aumentare la tendenza delle vie aeree superiori a collassare, portando all'insorgenza di apnee.

Farmaci miorilassanti come sedativi, ipnotici e oppiacei, la privazione del sonno e la posizione supina durante il sonno possono tutti aggravare la SAS Ostruttiva. Tuttavia, il grado di peggioramento dei disturbi respiratori del sonno può variare in base ai meccanismi fisiopatologici specifici e alla risposta fisiologica individuale.

Anche la riduzione della pervietà nasale dovuta a congestione o anomalie anatomiche, così come le allergie respiratorie, possono contribuire potenzialmente alla SAS Ostruttiva.⁽⁵⁾

In passato, il metodo più comunemente utilizzato per la diagnosi dell'Apnea Ostruttiva del Sonno è stata la polisonnografia notturna (PSG), come verrà descritta dettagliatamente nei capitoli successivi; tuttavia, sempre più spesso vengono adottate metodiche diagnostiche più semplici, che possono essere effettuate direttamente a domicilio, anziché in ambiente ospedaliero ad esempio: la PSG portatile di terzo livello (35). La registrazione video dei movimenti anomali durante il sonno può essere un utile strumento per identificare altri disturbi oltre alla SAS Ostruttiva. Tuttavia, per ottenere una valutazione accurata, è necessario eseguire un esame manuale della Polisonnografia, utilizzando linee guida specifiche per interpretare l'elettroencefalogramma (EEG) durante il sonno e assegnare punteggi agli eventi respiratori e ad altri eventi.

Negli ultimi tempi, stanno diventando sempre più popolari i sistemi di registrazione semplificati, noti come PSG respiratoria o poligrafia. Questi sistemi misurano vari parametri, tra cui il flusso d'aria alle vie aeree, gli sforzi respiratori, la saturazione di ossigeno e la frequenza cardiaca, ma non includono l'EEG. Questi sistemi presentano vantaggi come una maggiore disponibilità, costi inferiori, portabilità e maggiore

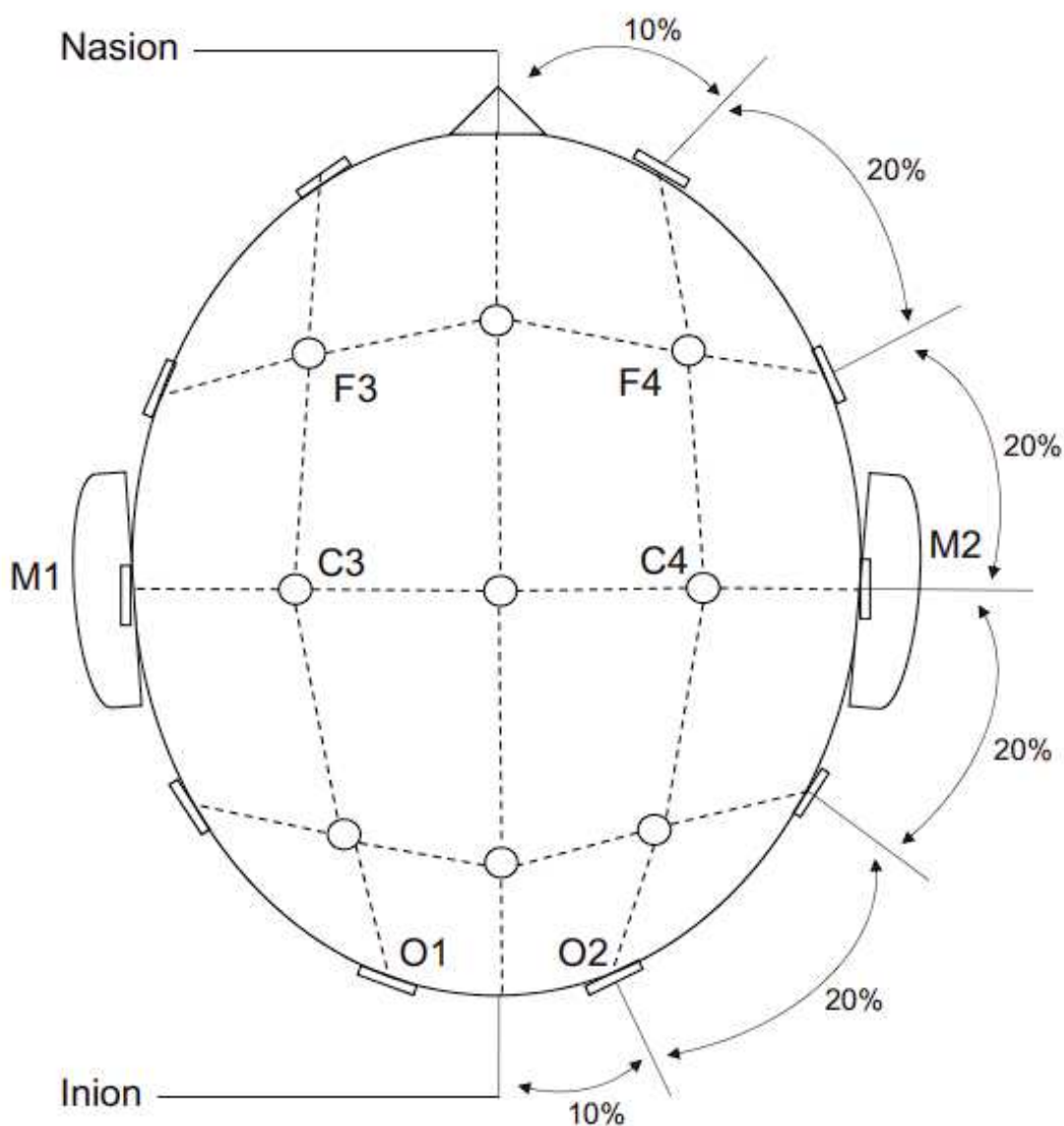


Figura 1.2: Esempio di disposizione degli elettrodi per EGG (1).

comodità per i pazienti, che possono utilizzarli direttamente a casa propria.

L'ossimetria notturna, che monitora la saturazione di ossigeno nel corso della notte, viene talvolta utilizzata come test di screening per individuare pazienti con SAS Ostruttiva. Tuttavia, è importante notare che ci sono limitazioni significative nell'uso dell'ossimetria da sola per la diagnosi di SAS Ostruttiva, poiché non fornisce informazioni complete sulla complessità dei disturbi respiratori durante il sonno. Pertanto, è spesso necessario combinare l'ossimetria con altre misurazioni per una diagnosi più precisa e completa della SAS Ostruttiva e di altri disturbi correlati al sonno.

1.2 Apnee centrali del sonno

Le apnee centrali del sonno (CSA) sono meno comuni rispetto alle apnee ostruttive del sonno, ma comunque degne di attenzione. Si stima che le CSA colpiscano circa il 10-20% di tutti i pazienti affetti da apnee notturne, ma la loro prevalenza può variare notevolmente a seconda del gruppo di popolazione preso in considerazione. Gli uomini, in genere, sono più a rischio rispetto alle donne, e l'incidenza aumenta con l'età. Sono stati riscontrati tassi più elevati di CSA in pazienti con malattie cardiache, insufficienza respiratoria cronica e disturbi del sistema nervoso centrale(11).

Le CSA sono un tipo di disturbo respiratorio che differentemente delle apnee ostruttive, non si ha un ostacolo fisico ad impedire il flusso d'aria, ma nelle CSA il problema principale è di natura neurologica. Durante un'apnea centrale, il cervello smette temporaneamente di inviare segnali ai muscoli responsabili della respirazione causando un temporaneo fallimento del pacemaker pontomedullare che genera il ritmo respiratorio. Ciò comporta la perdita dello sforzo ventilatorio, e se dura 10 secondi o più, viene definita CSA. Questa interruzione involontaria della respirazione può durare da alcuni secondi a minuti, causando ipossia (basso livello di ossigeno nel sangue) e disturbando il sonno. Le apnee centrali possono essere categorizzate come "primarie" quando non sono associate ad altre condizioni mediche, o "secondarie" quando sono causate da patologie preesistenti (12) (13).

Molti sono i fattori di rischio che possono contribuire allo sviluppo delle apnee centrali del sonno, tra cui:

- **Età avanzata:** Le CSA sono più comuni negli anziani a causa dei cambiamenti neurologici legati all'invecchiamento.
- **Malattie cardiache:** Pazienti con insufficienza cardiaca congestizia o altre malattie cardiovascolari hanno un rischio maggiore.
- **Uso di farmaci:** Alcuni farmaci, come gli oppioidi, possono sopprimere la respirazione e aumentare il rischio di CSA.

La diagnosi delle CSA richiede una valutazione specialistica. Le principali modalità diagnostiche includono:

- **Polisonnografia:** Durante questo esame, vengono monitorati vari parametri durante il sonno, tra cui l'attività cerebrale, il flusso d'aria, l'ossigenazione del sangue e i movimenti corporei. Ciò consente di individuare gli episodi di apnea centrale e di valutare la loro gravità. In caso di CSA, non si verificano segnali neurali inspiratori nel tronco cerebrale e i nervi che controllano i muscoli inspiratori toracici rimangono inattivi. Di conseguenza, nella polisonnografia, l'apnea

centrale è caratterizzata dall'assenza di flusso d'aria attraverso il naso o la bocca e dalla mancanza di movimenti toracoaddominali.

- **Monitoraggio domiciliare:** In alcuni casi, è possibile effettuare il monitoraggio delle apnee centrali del sonno a casa del paziente utilizzando dispositivi portatili specifici.
- **Valutazione clinica:** Il medico considererà anche la storia clinica del paziente, i sintomi e i fattori di rischio per stabilire una diagnosi accurata.

In conclusione, le apnee centrali del sonno rappresentano una sfida per la salute del sonno e la qualità della vita dei pazienti. L'epidemiologia, la definizione, i fattori di rischio e le modalità diagnostiche qui esaminate sono fondamentali per la comprensione e la gestione di questa condizione. Il riconoscimento precoce e il trattamento adeguato delle CSA possono migliorare notevolmente la salute e la qualità del sonno dei pazienti.

Capitolo 2

Stato dell'arte e scopo della tesi

In questo capitolo si esplorerà la metodologia attualmente utilizzata negli ospedali per identificare e diagnosticare le apnee notturne. Un focus particolare sarà riservato ai nuovi approcci basati sul deep-learning che offrono nuove possibilità di supporto alla pratica clinica corrente.

2.1 Metodologie diagnostiche validate clinicamente

L'identificazione delle SAS è un tema che negli ultimi anni sta attirando l'attenzione della ricerca. La valutazione della “sola” sintomatologia (ad esempio russamento, sonnolenza giornaliera...) è insufficiente per una corretta diagnosi nel 50% dei casi, anche quando condotta da medici esperti nello studio dei disturbi del sonno, un percorso diagnostico efficace richiede che il paziente si sottoponga a test specifici. I test diagnostici utilizzabili comprendono:

- **Questionari** che, in varia forma, valutano la presenza di russamento notturno e del grado di sonnolenza e affaticamento diurni. Inoltre alcuni raccolgono informazioni su ulteriori variabili che lasciano sospettare l'occorrenza di SAS ostruttiva. Tra queste variabili si annoverano la storia di ipertensione arteriosa, l'età, il sesso, la circonferenza del collo, il BMI e il gruppo etnico di provenienza. Tra i questionari validati ed utilizzati nella pratica clinica ci sono: Stanford sleepiness scale, Epworth sleepiness scale e Berlin Questionnaire (36). Tuttavia l'utilizzo di scale valutative ha delle limitazioni come: (i) variabilità intra ed inter-clinico. Infatti l'outcome e la valutazione risposte dipende in larga parte dall'esperienza del clinico e dei clinici; (ii) bias indotti da risposte errate da parte del paziente; (iii) status emotivo del paziente, infatti, la risposta alle domande da parte dell'esaminando può essere largamente influenzata dal livello di stanchezza del paziente o dall'agitazione al momento dell'esame.

- **Test strumentali per la valutazione della sonnolenza diurna** che, in letteratura e nella pratica clinica, sono scarsamente impiegati: (i) Multiple sleep latency test, (ii) Maintenance wakefulness test. Il primo è un test di un'intera giornata che consiste in cinque sonnellini programmati. Dopo la prima prova di sonnellino, ogni prova di sonnellino deve iniziare 2 ore dopo l'inizio della prova precedente. Questo test viene sempre eseguito dopo uno studio del sonno che misura la qualità e la durata del sonno. Durante ogni prova di sonnellino, il paziente si sdraia tranquillamente a letto e cerca di addormentarsi. Una volta spente le luci, il test misura il tempo necessario per addormentarsi. Se non ci si addormenta entro 20 minuti, la prova del sonno terminerà. Durante il test, una serie di sensori misurano se il paziente sta dormendo e valutano anche la fase del sonno corrente. Il secondo test ha un protocollo simile al primo e consiste in quattro prove del sonno con pause di due ore. A volte si effettua anche uno studio del sonno la notte prima del test. È fondamentale che la notte prima del test si riesca a dormire in modo normale e di qualità. In caso contrario, potrebbe essere necessario spostare il test a un altro giorno. Durante il test il paziente indossa una serie di sensori posizionati sulla testa, sul viso e sul mento come è possibile vedere in figura 2.1. I segnali acquisiti indicano quando il paziente è addormentato o sveglio durante il test. Per ogni prova, il paziente siede a letto con la schiena e la testa sostenute da un cuscino e l'obiettivo è quello di rimanere svegli il più a lungo possibile. Non è consentito fare altre cose per cercare di tenersi svegli. Il test termina se il soggetto che sta compiendo l'analisi non si addormenta entro 40 minuti.
- Il gold standard diagnostico è rappresentato dalla **polisonnografia**. L'American Thoracic Society, l'American College of Chest Physicians e l'Association of Professional Sleep Society hanno definito indicazioni e standard per l'esecuzione della polisonnografia, fatti propri da enti nazionali quali l'Associazione Italiana Pneumologi Ospedalieri e l'Associazione Italiana Medicina del Sonno (15). La polisonnografia è un esame di tipo cardio-respiratorio che evidenzia la presenza di apnee ostruttive, apnee miste, apnee centrali ed ipopnee. La polisonnografia è considerata un esame non invasivo (per definizione si considerano "non invasivi" gli esami che sfruttano strumenti che non penetrano le difese naturali dell'organismo umano) e viene eseguito monitorando durante il riposo notturno una serie di parametri fisiologici tra cui il livello di ossigeno nel sangue, la frequenza cardiaca e la respirazione. Per fornire questi parametri, lo strumento impiega svariati sensori alcuni dei quali si considerano intrusivi per i pazienti tra cui:
 1. Una fascia elastica toracica;

2. Un saturimetro sul dito che varia a seconda dell'età del paziente che si sottopone all'esame;
3. Una cannula nasale per la rilevazione del flusso respiratorio;
4. Un sensore di posizione (di solito presente nel polisonnografo).

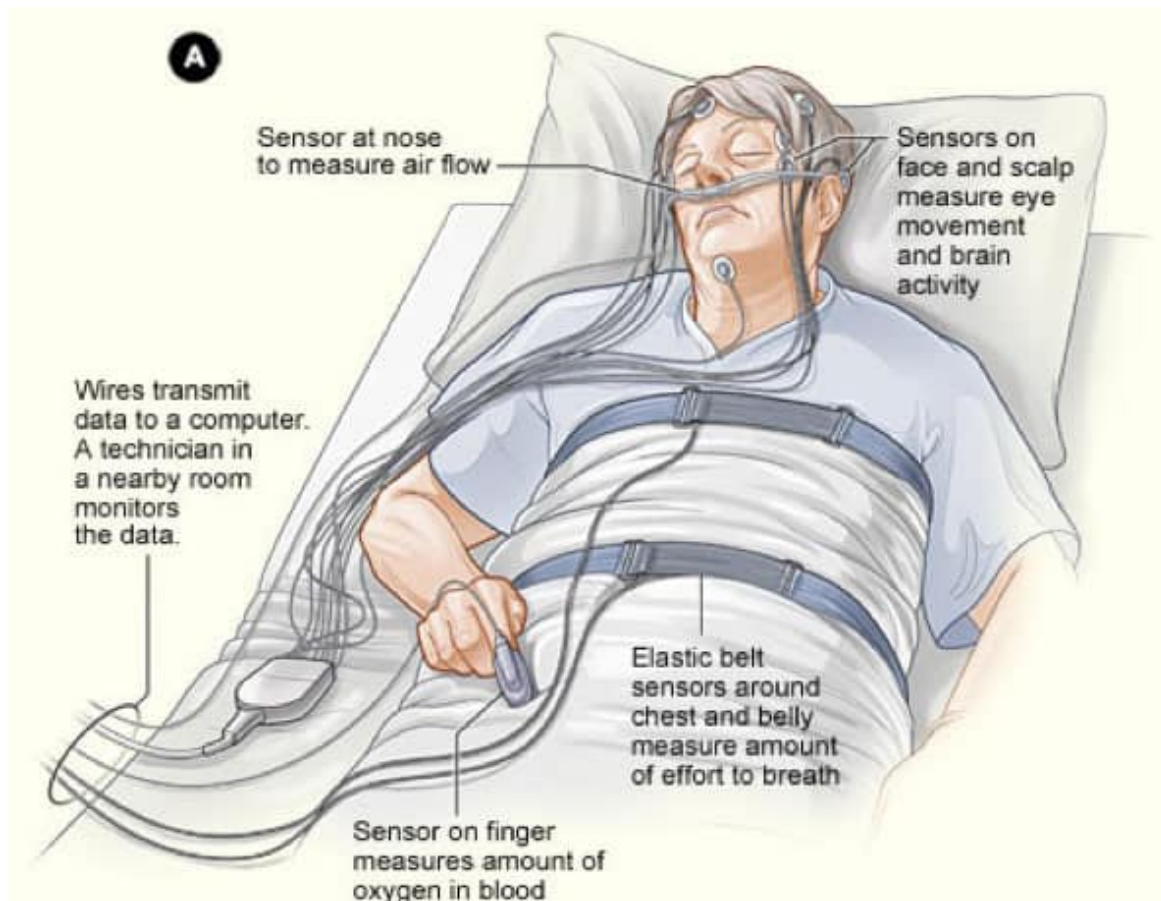


Figura 2.1: Disposizione della sensoristica durante un'analisi polisunnografica.

L'esecuzione della polisunnografia implica che il paziente debba obbligatoriamente dormire una o più notti presso il centro del sonno/ospedale per permettere allo strumento di rilevare e quantificare le alterazioni del ciclo sonno-veglia e quindi il deficit di sonno o l'eccesso di sonno di ciascun paziente (2). Tale procedura che, ad oggi, è considerata il gold standard clinico ha alcune limitazioni: (i) non può essere estesa alla valutazione in ambiente domiciliare, (ii) soprattutto quando l'indagine deve essere svolta su un paziente in età pediatrica la permanenza al centro clinico e la presenza della strumentazione possono indurre bias nella valutazione oltre che stress a livello psicofisico del bambino e del tutore (1)(16).

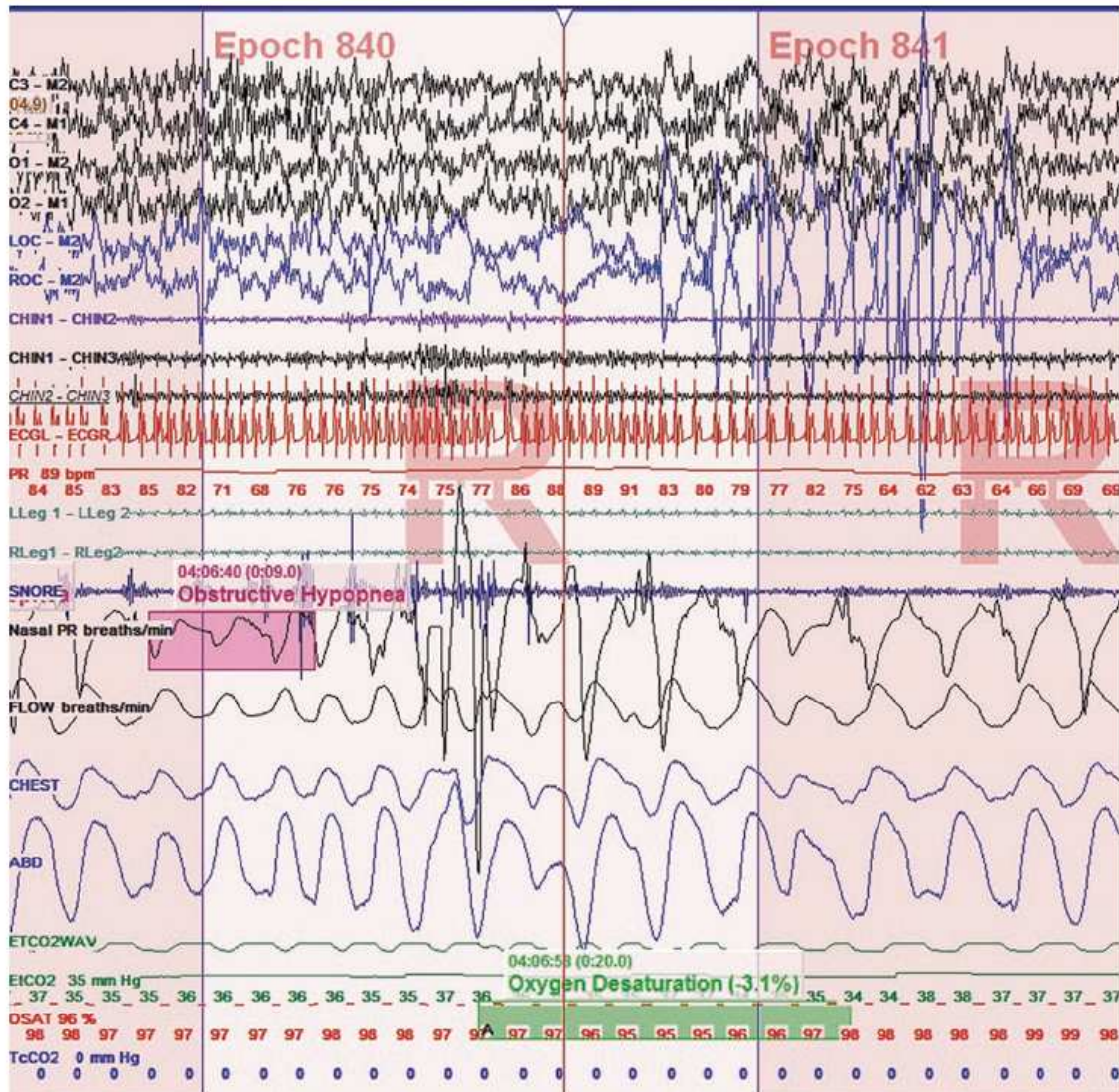


Figura 2.2: Esempio di schermata video relativa ad un'analisi polissonografica in presenza di un'apnea ostruttiva (2).

2.2 Scopo della tesi

In letteratura si rende più volte manifesta la necessità di proporre strumenti integrativi – possibilmente non intrusivi per il paziente – al gold standard diagnostico. Tali strumenti di supporto decisionale sono rilevanti per cogliere con anticipo evidenze della patologia e contrastare effetti negativi quali: (i) alta pressione sanguigna, (ii) problemi cardiaci o polmonari, (iii) crescita e sviluppo lenti soprattutto in pazienti adolescenti e per permettere ai clinici di implementare tempestivamente strategie terapeutiche atte a prevenire o ridurre le complicazioni legate alla SAS.

Con la volontà di supportare i clinici nel processo valutativo delle apnee notturne, il presente lavoro di tesi si pone l'obiettivo di identificare automaticamente, mediante

un approccio basato su deep learning, gli eventi apnoici di tipo centrale ed ostruttivo, basandosi esclusivamente sulle registrazioni dei cicli di sonno dei pazienti ottenute mediante l'utilizzo di microfoni tracheali. Tale approccio consente di acquisire dati in maniera notevolmente meno invasiva rispetto all'impiego della tradizionale sensoristica per la polisonnografia, rappresentando un passo avanti nel campo del monitoraggio delle apnee del sonno.

L'impatto di questa ricerca, oltre nell'impiego di sensori meno costosi ed invasivi per il paziente, risiede nella possibilità di automatizzare il processo di rilevamento degli eventi apnoici, riducendo la necessità di un approccio manuale per l'analisi di tutte le acquisizioni.

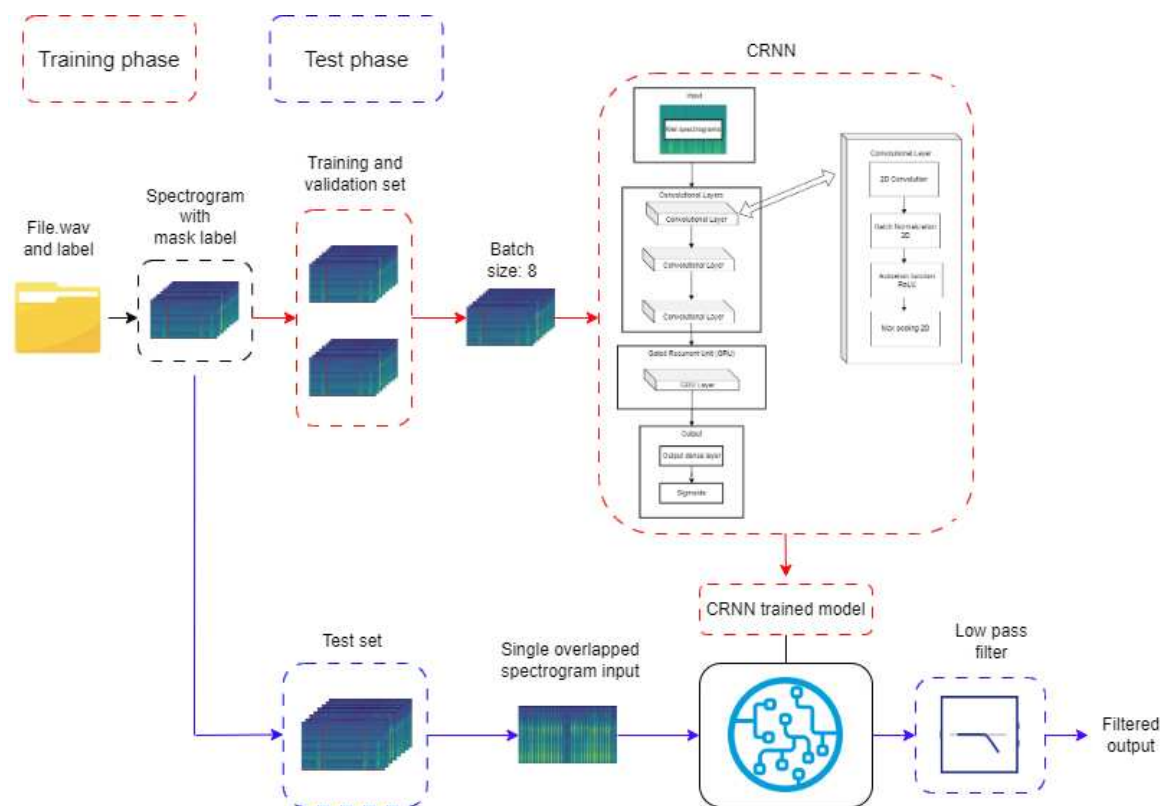


Figura 2.3: Diagramma di flusso del metodo proposto

2.3 Identificazione automatica di apnee notturne da audioregistrazioni

Di seguito si presentano contenuti recenti in letteratura che si concentrano sull'applicazione di metodologie di analisi automatica per l'identificazione di apnee notturne da audioregistrazioni.

- Korompili et al. (17), nel loro approccio, si basano sull'analisi audio delle registrazioni dei cicli di sonno realizzate mediante l'utilizzo di un microfono tracheale ed uno ambientale. Calcolano da prima la probabilità di respiro da brevi segmenti di 0.1 secondi estratti direttamente dai file audio del dataset. Questo calcolo è eseguito mediante un algoritmo di Voice Activity Detection (VAD), che identifica le porzioni del segnale audio associate all'attività vocale, che in questo caso è rappresentata dalla respirazione.

Successivamente, la probabilità media di respirazione viene calcolata su finestre consecutive di 0,1 secondi. Questa probabilità viene normalizzata per individuare le transizioni tra l'inalazione, l'esalazione e il silenzio. Per determinare quando avviene la transizione tra lo stato di respirazione e l'assenza di respirazione, viene stabilita una soglia basata sulla probabilità media di respirazione calcolata sugli ultimi 150 secondi del segnale audio. Inoltre, viene applicata un'altra soglia per eliminare le regioni identificate con una durata inferiore a 0,7 secondi, poiché sono considerate troppo brevi per essere considerate un respiro completo.

Per quanto riguarda il "rilevamento degli eventi apnoici", il risultato del rilevamento consiste in sequenze di zeri e uni che indicano la presenza o l'assenza di una respirazione. Per ogni segmento di respirazione individuato, viene calcolata la root mean square (RMS) come rappresentazione di quel segmento.

Per rilevare l'apnea, vengono selezionate le regioni in cui il valore medio RMS dei frame di respirazione rilevati è ridotto del 50% rispetto ai valori dei frame precedenti e successivi, e questa condizione deve perdurare per almeno 10 secondi. In altre parole, l'apnea è identificata quando si verifica un marcato cambiamento nella potenza del segnale audio, che dura almeno 10 secondi, indicando una pausa prolungata nella respirazione. Il dataset utilizzato nell'articolo sopra descritto è lo stesso che viene riportato nel lavoro di tesi.

- In (18) è stato sviluppato un modello di previsione focalizzato sulla rilevazione degli eventi apnoici utilizzando registrazioni dei cicli di sonno, ottenute da un microfono ambientale sfruttando una rete neurale convoluzionale.

Il processo di addestramento è stato condotto raccogliendo dati audio e dati provenienti dalla polisonnografia di diversi soggetti. I dati audio sono stati successivamente trasformati in spettrogrammi Mel, che sono stati utilizzati come input per la rete neurale convoluzionale. Questo spettrogramma Mel rappresenta una rappresentazione visiva delle caratteristiche audio nel dominio della frequenza.

L'intero processo di sviluppo del modello ha coinvolto una fase di addestramento iniziale su un gruppo di dati di addestramento. Successivamente, il modello è stato valutato in modo imparziale su un gruppo separato di dati di validazione per ottimizzare i parametri del modello. Durante la fase di inferenza, è stato utilizzato il metodo denominato "Detecting while Slicing" prima di valutare il modello su un gruppo di dati di test separato.

Per valutare l'efficacia del modello, è stato calcolato l'errore di previsione confrontando gli eventi rilevati dall'analisi audio con gli eventi rilevati dalla polisonnografia, che rappresenta il "gold standard" per la diagnosi degli eventi apnoici durante il sonno. Gli esami polisonnografici sono stati utilizzati quindi per la generazione delle label associate agli eventi di interesse.

- Un approccio simile a quello descritto in (18) è stato implementato anche da Nakano et al. (19) nel contesto dello studio, sono stati generati spettrogrammi a partire dalle registrazioni audio di cicli di sonno utilizzando finestre da 60 secondi comportando la creazione di un totale di 763.236 immagini. Inoltre sono state collezionate le analisi di polisonnografia di tutti i pazienti del dataset.

Per etichettare automaticamente queste immagini, è stato utilizzato un programma personalizzato per PC. Questo programma, basandosi su dati polisonnografici, ha assegnato automaticamente etichette a ciascun frame dell'immagine, identificando sia lo stato degli eventi respiratori sia lo stato di sonno/veglia.

Successivamente, sono stati addestrati due diversi modelli di Convolutional Neural Network (CNN) in modo indipendente. Il primo è stato addestrato per la classificazione dello stato degli eventi respiratori, cioè per identificare se si verificano o meno eventi respiratori in un determinato momento, mentre l'altro modello è stato addestrato per la classificazione dello stato di sonno/veglia, ovvero per identificare se il paziente è addormentato o sveglio in un dato momento.

- Prendendo spunto da recenti considerazioni che hanno mostrato le potenzialità del deep-learning rispetto alle tecniche standard di elaborazione del segnale Li et al. in (20) procedono come segue: in prima istanza è stato applicato un metodo di sottrazione spettrale con l'obiettivo di eliminare il maggior numero possibile di disturbi da rumore presenti nei segnali. Questo passaggio è fondamentale per garantire che l'analisi successiva sia basata su dati puliti.

Successivamente, è stato impiegato un metodo automatico e non supervisionato che ha combinato diversi algoritmi, tra cui il VerticalBox Algorithm, l'Analisi delle Frequenze e il Clustering Fuzzy C-means. Questa combinazione di algoritmi è stata utilizzata per segmentare gli episodi di russamento a partire dai segnali

audio precedentemente filtrati. riscontrando una buona capacità del metodo nell'identificare e isolare i momenti in cui si verifica il russamento nei segnali audio.

In seguito, sono state create delle Visual Geometry (VG) maps. Queste VG maps rappresentano graficamente i segnali audio utilizzando rappresentazioni vettoriali anziché rappresentazioni a pixel tradizionali. Queste immagini ottenute sono state utilizzate come input per una Convolutional Neural Network (CNN), l'uso di queste rappresentazioni vettoriali consente alla rete neurale di apprendere caratteristiche complesse dai segnali audio, contribuendo alla successiva fase di analisi e classificazione.

- Il sistema di classificazione del russamento proposto da Cheng et. al. (21) si compone principalmente di due fasi: l'estrazione delle caratteristiche del russamento e la costruzione del classificatore. Nella fase iniziale, vengono estratte le caratteristiche del russamento utilizzando Mel-frequency cepstral coefficients (MFCC), filter banks, Short-time Energy e linear predictive coding (LPC) dal frammento di russamento etichettato. Le quattro caratteristiche estratte vengono poi utilizzate come input per un modello Long Short-Term Memory (LSTM) al fine di realizzare la classificazione del russamento associato ad un evento apnoico.

L'utilizzo di approcci basati su deep learning è fondamentale per sviluppare modelli più robusti per l'identificazione degli eventi apnoici, considerando la complessità e la variabilità di tali eventi. Tuttavia, confrontare diversi approcci è spesso difficoltoso poiché la maggior parte dei casi non vengono forniti né i codici sorgente né i dataset utilizzati come in (21) (20) (19) (18) . Inoltre abbiamo anche una sostanziale differenza tra le metodologie che passano dallo sfruttare metodi statistici ad approcci misti o come (20) che per identificare le apnee si concentra sul rumore di russamento rendo il così un confronto diretto poco significativo.

L'articolo di Korompili et al. (17) sarà il punto di riferimento primario per questo elaborato di tesi, nonostante non faccia uso di metodologie di deep learning per l'analisi dei dati. Gli autori applicano il metodo VAD per analizzare i segnali audio di un dataset disponibile pubblicamente, utilizzato anche nelle analisi condotte e discusse nei capitoli successivi di questo elaborato. È importante sottolineare che il task svolto dagli autori, ovvero la classificazione binaria degli episodi di apnea (sia di tipo centrale che ostruttiva) rispetto a quelli che non rientrano nella categoria di apnea, è analogo a quello affrontato in questo elaborato di tesi.

Capitolo 3

Materiali e metodi

In questo capitolo, inizieremo con la presentazione del modello di rete neurale impiegato nel contesto della presente tesi. Successivamente, procederemo con una dettagliata esposizione del dataset utilizzato. Infine, saranno esposte e discusse le soluzioni implementate in questa tesi, le quali costituiranno oggetto di analisi nel capitolo successivo.

3.1 Architettura di rete neurale impiegata

Per lo sviluppo di questa tesi è deciso di utilizzare una Convolutional Recurrent Neural Network (CRNN), cioè una potente architettura di deep learning che combina efficacemente due componenti fondamentali: un blocco Convolutional Neural Network (CNN) e un blocco bidirectional Gated Recurrent Unit (BGRU). Questa struttura unica è stata progettata per risolvere compiti di classificazione complessi, e il suo successo è stato ampiamente documentato in numerosi articoli che ne giustificano l'implementazione (23) (24) (25), in particolare per questo task ci siamo ispirati all'articolo proposto da Emre et.al. (22).

Il blocco CNN è responsabile dell'estrazione delle caratteristiche dalle immagini di input. Utilizzando una serie di layer convoluzionali seguiti da layer di pooling, la CNN è in grado di catturare pattern spaziali di varia complessità. Queste caratteristiche estratte sono fondamentali per comprendere il contenuto delle immagini e rappresentano il primo passo nel processo di classificazione.

Il blocco BGRU, d'altro canto, agisce come classificatore finale. Questo tipo di rete ricorrente è stato scelto per la sua capacità di gestire sequenze di dati, rendendolo ideale per l'analisi di immagini e video. Il GRU è in grado di considerare le informazioni estratte dal CNN e di catturare le relazioni temporali delle sequenze di dati prese in input. Questa capacità di analizzare il contesto sequenziale è ciò che rende la rete CRNN così efficace in compiti di classificazione (26) (27).

Questa architettura CRNN è stata implementata utilizzando il framework PyTorch, il che ne consente una facile adattabilità e personalizzazione per varie applicazioni. La flessibilità di PyTorch offre un ambiente ideale per sperimentare con diverse configurazioni di rete e ottimizzare le prestazioni per specifici compiti di classificazione basati su immagini.

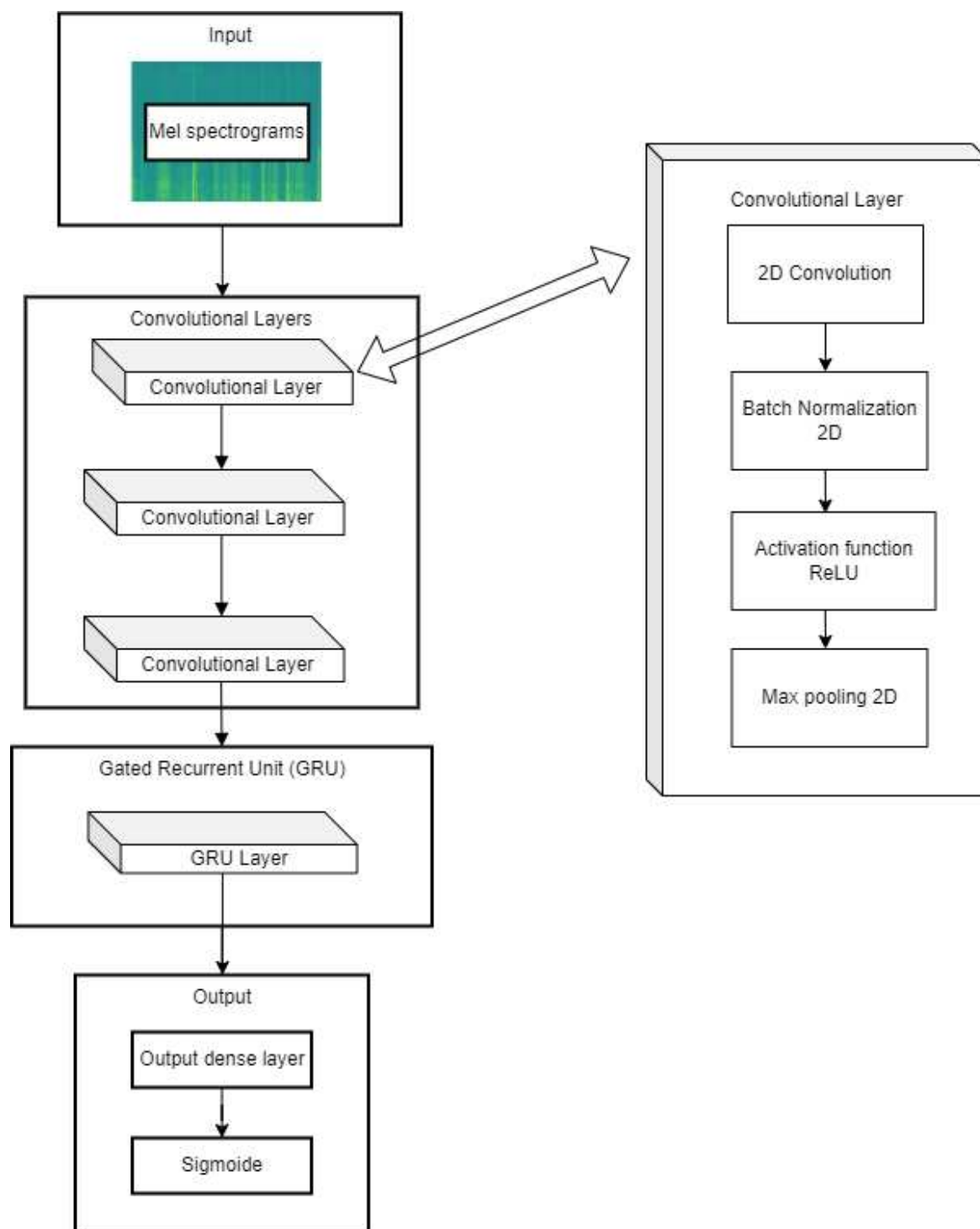


Figura 3.1: Diagramma a blocchi CRNN.

Come vediamo in figura la CNN è composta da 3 layer convuluzionali strutturati nel seguente modo:

- **Convoluzione 2D:** La convoluzione 2D¹ è un'operazione matematica che permette di trasformare un'immagine (o una matrice bidimensionale di dati) applicando un piccolo filtro (chiamato kernel) che scorre su di essa. In ogni posizione, si moltiplicano i valori sovrapposti del filtro e dell'immagine, poi si sommano questi risultati per ottenere un nuovo valore nell'immagine di output. Questa operazione è utilizzata per evidenziare caratteristiche specifiche nell'immagine, come contorni o pattern. È ampiamente utilizzata nell'elaborazione delle immagini e nell'apprendimento automatico per il riconoscimento di oggetti(?). Tra i parametri fondamentali di questa funzione troviamo:
 - **Padding:** Il padding² è una tecnica utilizzata per gestire la dimensione dell'output della convoluzione. Può essere impostato come "valid" (senza padding) cioè l'output risulta più piccolo rispetto all'input o "same" (con padding) Vengono aggiunti zeri intorno all'input in modo che l'output abbia le stesse dimensioni dell'input originale.
 - **Filter size:** Si riferisce alle dimensioni del filtro (kernel) utilizzato nella convoluzione. Ad esempio, in una convoluzione 2D 3x3, il filtro ha una dimensione di 3x3 pixel. La dimensione del filtro determina l'area locale dell'immagine esaminata o della feature map in ogni passaggio della convoluzione.
 - **Stride:** Lo stride esprime lo spostamento in numero di pixel che il filtro copia mentre scorre sull'input durante l'operazione di convoluzione. Uno stride più grande riduce la dimensione dell'output, mentre uno stride più piccolo produce un output più grande. Lo stride influisce sulla quantità di sovrapposizione tra le regioni di input analizzate dal filtro e sulla dimensionalità dell'output finale. Stride più piccoli possono conservare più dettagli spaziali.
- **Batch Normalization 2D:** La Batch Normalization 2D³ (BN2D) è una tecnica utilizzata nelle reti neurali per rendere l'allenamento più stabile e veloce. Essa normalizza i dati in modo che abbiano una media vicina a zero e una deviazione standard vicina a uno, riducendo così i problemi come l'instabilità dell'allenamento. Questo processo di normalizzazione avviene calcolando medie e deviazioni standard per ogni batch di dati durante l'allenamento e applicando

¹Url: <https://pytorch.org/docs/stable/generated/torch.nn.Conv2d.html>

²Url: <https://pytorch.org/docs/stable/generated/torch.nn.functional.pad.html>

³Url: <https://pytorch.org/docs/stable/generated/torch.nn.BatchNorm1d.html>

queste normalizzazioni ai dati in uscita dallo strato della rete neurale. In breve, la BN2D aiuta a migliorare l'efficienza e l'efficacia delle reti neurali durante l'allenamento (?).

- **ReLU:** La Rectified Linear Unit (ReLU)⁴ è una funzione di attivazione comunemente utilizzata nelle reti neurali. La sua operazione è molto semplice: se il valore in ingresso è positivo, lo restituisce invariato; se è negativo, restituisce zero. La ReLU è ampiamente apprezzata per la sua semplicità e il suo contributo nel mitigare il problema della scomparsa gradiente (Vanishing gradient (34)) durante l'allenamento delle reti neurali. In breve, la ReLU aggiunge non linearità al modello, attivando solo i valori positivi e lasciando gli altri a zero.
- **Max Pooling 2D:** Il Max Pooling 2D⁵ è una tecnica utilizzata nell'elaborazione delle immagini per ridurre la dimensione dei dati e mantenere le caratteristiche più rilevanti di un'immagine. Funziona suddividendo un'immagine in piccoli riquadri e selezionando il valore del pixel più alto all'interno di ciascun riquadro. Questo valore massimo diventa quindi il rappresentante di quella regione. In sostanza, il Max Pooling 2D semplifica l'immagine, mantenendo le informazioni più importanti, come i bordi o le caratteristiche distintive, e riducendo la dimensione dei dati, il che può rendere più efficiente il processo di elaborazione dell'immagine o il riconoscimento di pattern.

I parametri caratterizzati dei tre layer convoluzionali sono riassunti nella seguente tabella:

Layers	Number of filter	Filter size	Stride	Padding	Pool Size
Conv1	96	5x5	1	Adaptive	5x2
Conv2	128	5x5	1	Adaptive	5x2
Conv3	128	5x5	1	Adaptive	3x2

Tabella 3.1: Tabella riassuntiva dei parametri nei layer convoluzionali. Con *Adaptive* si intende che il padding viene scelto autonomamente dall'algoritmo al fine di mantenere invariate le dimensioni delle immagini a seguito di una convoluzione

3.2 Dataset

Nel contesto di questa tesi, l'articolo di riferimento del dataset utilizzato è quello proposto in (3). Questo articolo presenta 279 analisi di polisonnografia con registrazioni sincronizzate del suono tracheale e ambientale in formato open-access. L'obiettivo

⁴Url: <https://pytorch.org/docs/stable/generated/torch.nn.ReLU.html>

⁵Url: <https://pytorch.org/docs/stable/generated/torch.nn.MaxPool2d.html>

principale di questo dataset è supportare studi comparativi sull'efficacia dei microfoni, sulla rilevazione delle apnee basata su caratteristiche ricorrenti e sull'aumento dei dati per la ricerca correlata alla polisonnografia.

Gli autori forniscono dettagli sulla metodologia utilizzata per raccogliere il dataset, discutono considerazioni etiche e presentano statistiche essenziali. Sottolineano in particolare l'importanza di avere dataset bilanciati e ampi per lo sviluppo accurato di sistemi di rilevazione della sindrome delle apnee notturne.

I dati per questo studio sono stati raccolti presso l'Unità di Studio del Sonno dell'Ospedale Generale Sismanoglio - Amalia Fleming di Atene. Il processo di raccolta dati ha seguito un protocollo standard per la polisonnografia suddivisa, con la prima parte che coinvolge una polisonnografia diagnostica della durata approssimativa di quattro 4 ore. La seconda parte della polisonnografia è dedicata alla taratura del livello ottimale di pressione positiva continua delle vie aeree (CPAP) per eliminare gli eventi di apnea (14). Poiché questo dataset è stato progettato per lo studio del sonno senza intervento CPAP, sono stati inclusi solo i dati della prima parte dello studio di polisonnografia.

Il monitoraggio dei canali della polisonnografia e l'acquisizione dei segnali sono stati effettuati utilizzando il software Sleepware G3.

Simultaneamente allo studio polisonnografico, sono stati registrati segnali audio utilizzando un registratore multitraccia portatile a doppio canale (Tascam DR-680 MK II) dotato di due microfoni di alta qualità:

- Un microfono a contatto Clockaudio CTH100 (figure 3.3 3.4) posizionato sulla trachea del paziente. Questo microfono è stato appositamente progettato per rilevare le vibrazioni del collo ed era altamente insensibile al rumore ambientale. Ha un'impedenza di ingresso di 900Ω e una banda passante compresa tra 350 Hz e 8 kHz.
- Un microfono a condensatore per misure ultralineare (Behringer ECM8000) posizionato a circa 1 metro sopra il letto del paziente, nella zona della testa. Questo microfono era omnidirezionale, con un'impedenza di ingresso di 200Ω e una vasta risposta in frequenza compresa tra 15 Hz e 20 kHz.

Entrambi i segnali audio sono stati campionati a una frequenza di 48 kHz e memorizzati su una scheda SD in formato audio Waveform Audio Format (.wav) a 24 bit non compresso. Queste registrazioni audio hanno integrato i dati della polisonnografia ed sono essenziali per l'analisi dei suoni della respirazione durante gli studi del sonno. Nel nostro studio verranno considerati esclusivamente di i dati provenienti dal microfono a contatto. Un esempio del setup della postazione utilizzata per la raccolta dati è mostrato in figura 3.2.



Figura 3.2: Setup della postazione usata per l'acquisizione dei dati (3).

Ogni studio di polisonnografia condotto viene sottoposto a una dettagliata caratterizzazione medica. Le fasi del sonno e gli eventi di apnea per ciascun paziente vengono valutati da due specialisti. Il processo di valutazione prevede che un tecnico certificato effettui la valutazione iniziale, seguita da una valutazione finale eseguita da un medico altamente esperto e certificato. Questa valutazione finale include la verifica degli eventi annotati veri positivi e l'aggiunta di eventuali eventi mancanti. Sebbene il processo di valutazione non sia cieco, il protocollo seguito è progettato per migliorare l'accuratezza dell'identificazione.

L'accordo tra gli osservatori è stato valutato per garantire un'alta precisione nell'annotazione degli eventi. Inoltre, è stato sviluppato un algoritmo per quantificare la diminuzione dell'ampiezza del flusso d'aria all'interno di ciascun evento valutato, fornendo una convalida dei dati e migliorando ulteriormente l'accuratezza nell'identificazione degli eventi veri positivi.

La rilevazione degli eventi apnoici durante le ore di sonno registrate viene eseguita manualmente osservando contemporaneamente tutti i canali del sistema di polisonnografia, seguendo i criteri stabiliti per la valutazione degli episodi di apnea. È importante notare che le registrazioni audio non vengono considerate nel processo di diagnosi.



Figura 3.3: Microfono a contatto Clockaudio CTH100.



Figura 3.4: Esempio di utilizzo del microfono a contatto Clockaudio CTH100.

Uno dei passaggi critici in questo studio è la sincronizzazione dei segnali audio acquisiti con i segnali polisonnografici, ottenuti da sistemi attivati separatamente. L'algoritmo di sincronizzazione utilizza il suono tracheale registrato da un microfono di contatto separato e di qualità inferiore all'interno del sistema polisonnografici, che ha una frequenza di campionamento di 500Hz e una profondità di bit di 16 bit.

Il processo di sincronizzazione prevede i seguenti passaggi:

- **Adeguamento delle Frequenze di Campionamento:** Inizialmente, è necessario ridurre la frequenza di campionamento del segnale audio di alta qualità del suono tracheale per farla corrispondere alla frequenza di campionamento di 500Hz del canale "Russare" del polisonnografo.
- **Estrazione dell'Inviluppo del Segnale:** Viene estratto l'inviluppo del segnale sia del segnale audio di alta qualità del suono tracheale che del segnale "Russare" del polisonnografo utilizzando la trasformata di Hilbert (31) (32).
- **Analisi di Correlazione Incrociata:** Viene calcolata la correlazione incrociata dei due inviluppi dei segnali. Questa analisi aiuta a stimare il ritardo nell'attivazione dei due sistemi identificando il valore di differenza temporale che massimizza la correlazione incrociata.

Per garantire una sincronizzazione accurata, viene esaminato un minimo di 10 minuti di segnali. Inoltre, ai supervisori dello studio del sonno viene dato l'incarico

di attivare i due sistemi con il minor ritardo possibile, di solito entro 30 secondi. Le polisonnografie con un errore che supera i 2 secondi in questo passaggio vengono escluse.

La memorizzazione dei dati della polisonnografia e dei dati del suono della respirazione è stata effettuata utilizzando il formato dati europeo (EDF), comunemente utilizzato per la memorizzazione e il trasferimento di dati medici nel campo degli studi del sonno e dell'assistenza sanitaria.

I dati audio aggiuntivi sono stati ottenuti da file WAV non compressi ed integrati nei file EDF con i nomi dei canali "label" e "Microphone", corrispondenti al segnale del suono tracheale e al segnale del microfono ambientale, rispettivamente. Per gestire i problemi legati alla sincronizzazione, è stata utilizzata la politica di riempimento con zeri per le parti mancanti dei segnali audio.

È importante notare che l'EDF ha limitazioni in termini di profondità di bit rispetto ai file WAV originali, in quanto l'EDF comunemente utilizza una profondità di 16 bit. La conversione dalle registrazioni audio iniziali di alta qualità a 24 bit a una profondità di 16 bit è stata ottenuta scartando i bit meno significativi di ciascun valore di campione. L'errore assoluto statistico introdotto da questa conversione ha una media di 7.6291×10^{-6} , con un valore massimo che non supera $1,5259 \times 10^{-5}$. Questi valori comportano una riduzione minima della qualità nelle registrazioni audio e non dovrebbero influire significativamente sugli studi relativi alle proprietà dei suoni della respirazione o dei russamenti.

Nel dataset, ulteriori file .rml sono responsabili di tutte le annotazioni relative a ciascuno studio del paziente. Questi file contengono informazioni sulle fasi del sonno ed eventi etichettati. Gli eventi annotati sono categorizzati per famiglia, tra cui "respiratori", "neurologici", "legati all'attività degli arti", "nasali" e "cardiaci". In particolare, gli episodi "respiratori", che sono il focus principale di questo rapporto, comprendono vari episodi correlati all'apnea, come "Apnea Ostruttiva", "Apnea Centrale", "Apnea Mista" e "Ipopnea". Inoltre, i file .rml includono annotazioni per eventi relativi alla desaturazione dell'ossigeno e eventi di risveglio. Per un utilizzo più agevole delle annotazioni sono stati creati dei file .json per ogni paziente dove sono contenuti solamente gli eventi respiratori di apnea con le relative tipologie.

I dati di ciascun paziente nel dataset sono divisi in parti della durata di 1 ora e successivamente ogni ora è stata segmentata in porzioni da 5 min così da facilitarne la gestione e l'analisi. Ciascun candidato è identificato da un numero di paziente rappresentativo unico nell'intervallo da 995 a 1701.

Per quanto riguarda lo sviluppo della tesi sono stati usati esclusivamente i file audio tracheali contrassegnati come "label" ad i file .json contenenti le label degli eventi di interesse.

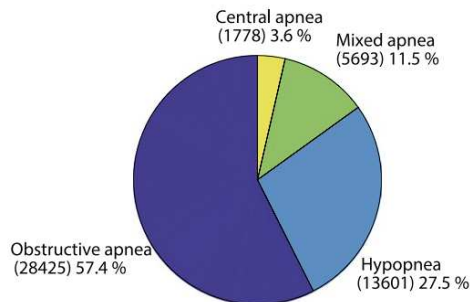


Figura 3.5: Distribuzione dei tipi di apnea all'interno del dataset (3).

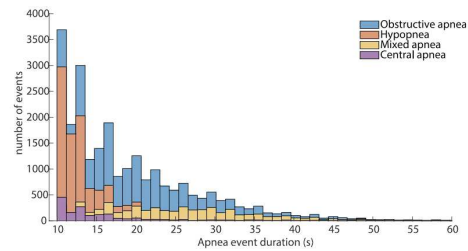


Figura 3.6: Durata in s degli eventi annotati nel dataset e relativa numerosità (3).

3.3 Pre-processing del dataset

Al fine di conferire al presente studio un aspetto congruente con l'ambiente clinico autentico, si è deciso di procedere con la suddivisione del dataset in sottoinsiemi destinati all'addestramento, alla validazione e alla verifica, privilegiando la stratificazione in base al numero di pazienti anziché al conteggio totale degli eventi apnoici. La distribuzione dei dati è stata organizzata secondo il seguente schema: l'80% dei pazienti è stato assegnato al set di addestramento (training set), mentre il restante 20% è stato ulteriormente suddiviso equamente in due parti, costituendo il set di validazione (validation set) e il set di test (test set), entrambi equivalenti al 10% del totale dei pazienti.

Prima di procedere con le fasi di addestramento, è stata eseguita un'analisi preliminare del dataset con lo scopo di identificare e gestire eventuali problematiche presenti. Inizialmente, si è effettuato un controllo di coerenza sulle registrazioni di 5 minuti, verificandone l'uniformità della lunghezza, della frequenza di campionamento e del numero di canali. Le registrazioni non conformi a tali criteri sono state etichettate come corrotte e, conseguentemente, escluse dall'analisi.

Successivamente, si è condotta un'analisi delle etichette associate al dataset al fine di comprenderne la distribuzione di probabilità degli eventi all'interno del campione e visualizzare la durata media di tali eventi.

Come si evince chiaramente dalla figura 3.7, la durata media degli eventi risulta essere approssimativamente pari a 12 secondi. Ciò implica che, al fine di ottenere un quadro accurato e completo della manifestazione delle apnee, è necessario lavorare con contesti temporali di almeno 20 secondi. In questa prospettiva, per l'addestramento del modello è stata adottata una finestra temporale di 50 secondi. Tale scelta metodologica è stata compiuta con l'obiettivo di addestrare la rete neurale su un contesto temporale più esteso, così da contestualizzare l'evento ed includere caratteristiche significative per il riconoscimento.

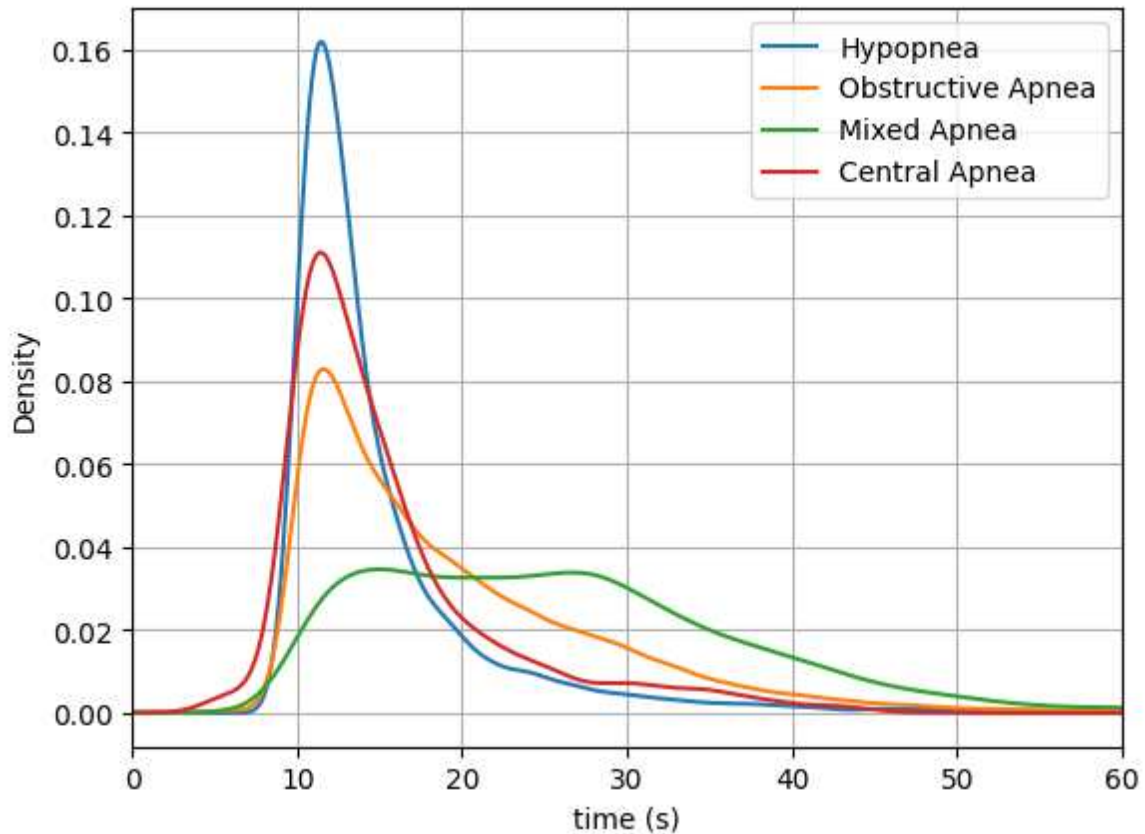


Figura 3.7: Grafico densità di probabilità-durata degli eventi annotati.

A causa delle dimensioni considerevoli del dataset e della diversità significativa degli eventi presenti, è stata adottata una strategia per semplificarne la gestione. In particolare, sono stati creati tre data-frame estesi a partire dai file .json contenenti le etichette degli eventi relativi al training-set, al validation-set e al test-set.

I file .json forniscono inizialmente solo quattro informazioni per ogni evento annotato, ossia:

- **Family:** che indica l'appartenenza dell'evento alla famiglia degli eventi respiratori.
- **Type:** che specifica la tipologia dell'evento respiratorio.
- **Start:** che rappresenta il momento di inizio dell'evento in termini di secondi.
- **End:** che indica il momento di conclusione dell'evento in termini di secondi.

Nella fase di costruzione del nuovo file di metadati (data-frame esteso) come possiamo vedere in figura 3.2, tutte le informazioni contenute nei file .json sono state conservate e ulteriormente arricchite con nuovi dettagli, tra cui:

- **Patient ID:** che rappresenta l'identificativo numerico del paziente a cui è associata l'apnea.
- **Relative Start:** che corrisponde al timestamp in secondi dell'inizio dell'evento, riferito all'interno di un intervallo di 300 secondi.
- **Relative End:** che denota il timestamp in secondi della conclusione dell'evento, sempre entro il medesimo intervallo di 300 secondi.
- **File Subcategory:** che rappresenta il numero di riferimento della sottodirectory secondo la quale il dataset è organizzato.
- **File Category:** che indica il numero di riferimento della directory principale secondo la quale il dataset è organizzato.
- **Duration:** che esprime la durata dell'evento in termini di differenza tra l'istante di fine e quello di inizio (End - Start).
- **File Name:** che costituisce il percorso del file.

Successivamente, a partire dal nuovo dataframe, sono state create delle maschere, che sono essenzialmente vettori binari composti da valori uno e zero. Queste maschere come mostrato in figura 3.8 sono state generate per l'individuazione degli eventi all'interno di ciascun file registrato nel dataframe. In termini più semplici, ciò che è stato realizzato consiste nell'assegnare sequenze di valori uno nelle posizioni corrispondenti alle etichette che indicano la presenza di un evento di apnea, mentre le altre posizioni sono state riempite con valori zero. Questo elemento risulterà poi indispensabile per la generazione delle label per la fase di addestramento.

Un esempio di questa operazione è illustrato nella figura 3.8.

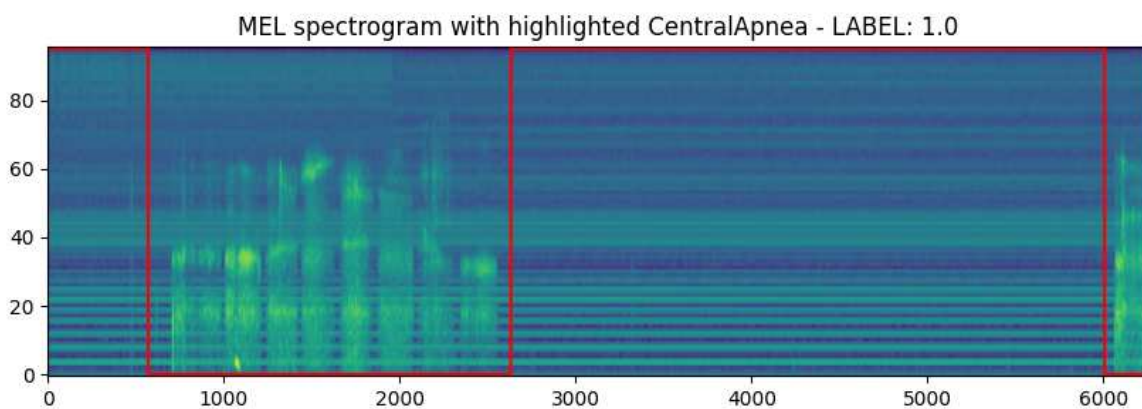


Figura 3.8: In rosso la maschera generata per identificare le apnee.

patient id	family	type	start	end	relative start	relative end	duration	file sub cat	file cat
1621	Respiratory	ObstructiveApnea	9981.0	9998.5	81.0	98.5	17.5	3	9
1631	Respiratory	CentralApnea	21767.0	21781.0	167.0	181.0	14.0	7	0
1656	Respiratory	Hypopnea	16432.5	16454.5	232.5	254.5	22.0	5	6
1623	Respiratory	ObstructiveApnea	992.5	1018.0	92.5	118.0	25.5	1	3

Tabella 3.2: Dataframe esteso.

Questo approccio alla gestione dei dati è stato adottato con l'obiettivo di semplificare la comprensione e la manipolazione del dataset, tenendo conto delle sue dimensioni notevoli e della complessità delle informazioni relative agli eventi annotati.

Nella seguente sezione, procederemo a descrivere in dettaglio la classe creata per alimentare il processo di addestramento della rete neurale, nonché gli iperparametri utilizzati e le metriche valutative scelte per quantificare l'efficacia del modello, le procedure di testing e gli esperimenti svolti.

3.4 Creazione del dataset di addestramento e impostazioni di addestramento

Per iniziare, è fondamentale esaminare il processo di creazione della classe dedicata alla generazione del dataset e come si è affrontato il problema della grande disuniformità tra eventi positivi (apnea) ed eventi negativi (non apnea) così da ottenere un dataset bilanciato per l'addestramento.

La classe in questione è stata denominata *MyDataset* ed è stata progettata per la creazione di un dataset personalizzato per l'addestramento della rete neurale precedentemente descritta. La classe è scritta in Python utilizzando il *framework PyTorch* ed è composta da vari metodi e funzionalità per la gestione dei dati.

La classe *MyDataset* accetta diversi parametri nel suo costruttore:

- **meta_df**: Un dataframe contenente metadati sui dati, come il nome del file e il tipo di apnea.
- **data_path**: Il percorso alla directory in cui sono archiviati i dati audio.
- **win_len**: La lunghezza della finestra utilizzata per estrarre i campioni in secondi.
- **thr**: La soglia per il calcolo della label associata al campione.
- **force_balance**: Il numero di cicli per bilanciare il dataset.
- **apnea_type**: Il tipo di apnea desiderato (default: "all").

La classe inizia con la selezione dei dati in base al tipo di apnea specificato, se non è "all", e estrae i nomi dei file path unici dai metadati. Successivamente, viene inizializzata una classe per la generazione degli spettrogrammi Mel dalla libreria Torchaudio con le specifiche per la trasformazione dello spettrogramma.

La classe offre anche un metodo `resample_mask` per ri-campionare una maschera di apnea al fine di adattarla alla lunghezza dello spettrogramma. Il metodo `__len__` restituisce la lunghezza del dataset, che è uguale al numero di file path unici delle registrazioni audio, mentre il metodo `__getitem__` è utilizzato per ottenere un elemento dal dataset.

All'interno del metodo `__getitem__`, vengono caricati i dati audio e le maschere corrispondenti dal file, quindi viene calcolato lo spettrogramma del file audio completo. Successivamente, lo spettrogramma viene normalizzato in scala logaritmica e quindi riportato nel range $[0, 1]$. La maschera viene ri-campionata per adattarla alla lunghezza dello spettrogramma.

Infine, il metodo cerca una finestra positiva e una finestra negativa per bilanciare il dataset. Le finestre positive sono determinate da una somma degli elementi della maschera superiore alla soglia, mentre le finestre negative sono determinate da una somma degli elementi della maschera inferiore alla soglia. Entrambe le finestre vengono restituite insieme alle rispettive etichette binarie. L'importanza di avere un dataset bilanciato in fase addestramento consente di:

- **Evitare il bias del modello:** Un dataset sbilanciato, cioè un dataset in cui una classe ha molte più istanze rispetto alle altre, può portare il modello a sviluppare un bias verso la classe maggioritaria. Il modello tenderà a predire più spesso la classe maggioritaria a scapito delle altre, riducendo l'accuratezza complessiva del classificatore.
- **Migliorare le prestazioni complessive:** Un dataset bilanciato consente al modello di apprendere in modo più equilibrato da tutte le classi. Questo può portare a prestazioni più elevate sulle classi minoritarie, che potrebbero essere più importanti dal punto di vista pratico o avere un impatto maggiore nella situazione reale.
- **Ridurre l'overfitting:** Quando si addestra un modello su un dataset sbilanciato, il modello può soffrire di overfitting sulla classe maggioritaria. In altre parole, il modello potrebbe apprendere troppo dai dati della classe maggioritaria e avere difficoltà a generalizzare correttamente sui dati di test o su nuovi dati.

- **Rappresentare i dati:** Un dataset bilanciato è più rappresentativo della distribuzione reale delle classi nel mondo reale. Questo aiuta il modello a catturare meglio la variabilità delle diverse classi e a fare previsioni più accurate.

Per quanto riguarda la validazione dei dati in fase di addestramento per la generazione del dataset ci si affida sempre alla classe `MyDataset` definita sopra cambiando però la directory di lavoro ed il relativo file di metadati con quella in cui sono contenuti i dati dei pazienti per la validation.

Il suo approccio alla selezione delle finestre di dati positive e negative è progettato per bilanciare il dataset, il che è spesso cruciale per la buona performance di un modello di apprendimento automatico. Certamente, è fondamentale anche fare un piccolo richiamo gli iperparametri tipici di qualsiasi classificatore. Tra i principali emergono i seguenti:

- **Learning rate:** Il learning rate è un parametro importante nell'addestramento dei modelli di machine learning e deep learning, poiché controlla la dimensione dei passi fatti durante il processo di ottimizzazione dei pesi. Un learning rate troppo alto può causare instabilità, mentre uno troppo basso può rendere l'addestramento lento. Nel nostro caso il learning rate viene gestito dallo `scheduler COSINEANNEALINGLR`⁶ (?), questo ci aiuta a superare minimi locali e accelerare la convergenza verso il minimo globale della funzione di Loss.
- **Batch size:** E' un parametro fondamentale nell'addestramento delle reti neurali e definisce quanti esempi di addestramento vengono utilizzati in ogni iterazione di un'epoca durante il processo di apprendimento automatico.
- **Numero epoche:** Rappresenta il numero di volte che l'intero dataset di addestramento viene passato attraverso la rete neurale durante il processo di apprendimento. La scelta del numero di epoche è cruciale per ottenere un modello addestrato in modo efficace e generalizzato.

3.5 Metriche valutative utilizzate

Nel campo della classificazione, l'accuratezza, la recall e la precision sono metriche valutative fondamentali che consentono di misurare l'efficacia di un modello o di un sistema nell'identificare correttamente gli elementi di interesse all'interno di un insieme di dati. Queste metriche forniscono una panoramica chiara delle performance del modello e sono spesso utilizzate per valutare l'adeguatezza di un sistema nell'ambito

⁶Url: https://pytorch.org/docs/stable/generated/torch.optim.lr_scheduler.CosineAnnealingLR.html

di applicazioni mediche, come il rilevamento di tumori, malattie o altre condizioni patologiche.

3.5.1 Accuratezza

L'accuratezza⁷ è una delle metriche più semplici e intuitive per valutare la performance di un modello. Essa rappresenta la frazione di predizioni corrette rispetto al numero totale di predizioni effettuate dal sistema. Formalmente, l'accuratezza può essere calcolata come:

$$\text{Accuratezza} = \frac{\text{Numero di predizioni corrette}}{\text{Numero totale di predizioni}} \quad (3.1)$$

L'accuratezza fornisce una stima generale della capacità di un modello di individuare correttamente gli elementi di interesse, ma può risultare fuorviante in contesti in cui le classi da individuare sono sbilanciate. Ad esempio, se stiamo cercando di individuare rari casi di una malattia, un modello che predice sempre la classe più frequente otterrebbe un'accuratezza elevata, ma sarebbe inutile nella pratica clinica (28).

3.5.2 Recall

La recall⁸ è una metrica che si concentra sulla capacità di un modello di individuare tutti i veri positivi all'interno del dataset. In altre parole, la recall misura la frazione di veri positivi rispetto al numero totale di veri positivi e falsi negativi, ed è particolarmente importante nelle applicazioni mediche in cui il costo di un falso negativo (ossia, non riconoscere una condizione clinica quando è presente) è elevato. La recall viene calcolata come:

$$\text{Recall}_i = \frac{\text{True positive}_i}{\text{True positive}_i + \text{False negative}_i} \quad (3.2)$$

Questa è una metrica cruciale nei task di individuazione clinica perché ci si aspetta che il sistema sia in grado di catturare il maggior numero possibile di casi positivi. Un elevato valore di recall indica che il modello è efficace nel riconoscere le condizioni patologiche, riducendo il rischio di falsi negativi che potrebbero avere gravi conseguenze per il paziente.

⁷Url: <https://torchmetrics.readthedocs.io/en/stable/classification/accuracy.html>

⁸Url: <https://torchmetrics.readthedocs.io/en/stable/classification/recall.html>

3.5.3 Precision

La precision⁹ è una metrica che misura la frazione di veri positivi rispetto al numero totale di predizioni positive effettuate dal modello. In altre parole, la precisione valuta quanto sia affidabile il modello quando emette una predizione positiva. La precisione si calcola come:

$$\text{Precision}_i = \frac{\text{True positive}_i}{\text{True positive}_i + \text{False positive}_i} \quad (3.3)$$

La precisione è importante perché fornisce un'indicazione della capacità del modello di evitare falsi allarmi. In situazioni in cui gli errori possono avere conseguenze gravi o costose, come nel caso di procedure mediche invasive, una precisione elevata è essenziale.

3.5.4 Precision-Recall Curve

La curva Precision-Recall¹⁰ è solitamente ottenuta variando la soglia di decisione del modello e calcolando la precision e la recall corrispondenti a ciascuna soglia. Quando questi valori vengono rappresentati in un grafico bidimensionale, otteniamo la curva Precision-Recall un esempio è mostrato in figura 3.9. Tipicamente, la curva inizia nell'angolo in basso a sinistra e si sposta verso l'angolo in alto a destra al diminuire della soglia di decisione (29) (30).

Nel contesto delle applicazioni mediche, la recall assume spesso un ruolo di primaria importanza rispetto all'accuratezza e alla precision. Questo perché, in molte situazioni, è cruciale identificare il maggior numero possibile di casi positivi, anche a costo di avere un numero maggiore di falsi positivi. Ad esempio, nel rilevamento precoce di una condizione patologica come le apnee, è fondamentale minimizzare i falsi negativi (ossia, non riconoscere un'apnea quando è presente) anche se ciò comporta un aumento dei falsi positivi (ossia, indicare la presenza di un'apnea anche quando non è presente).

Un sistema caratterizzato da un alta recall ma una bassa precision tende a produrre numerosi risultati, ma la maggior parte delle etichette previste risultano errate rispetto alle etichette di addestramento. In contrapposizione, un sistema con alta precision ma bassa recall restituisce un numero limitato di risultati, ma la maggior parte delle etichette previste sono corrette rispetto alle etichette di addestramento. L'ideale è rappresentato da un sistema che coniughi sia un'elevata precision che un'alta recall, restituendo numerosi risultati e garantendo che tutti essi siano etichettati correttamente (?).

⁹Url: <https://torchmetrics.readthedocs.io/en/stable/classification/precision.html>

¹⁰Url: https://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html

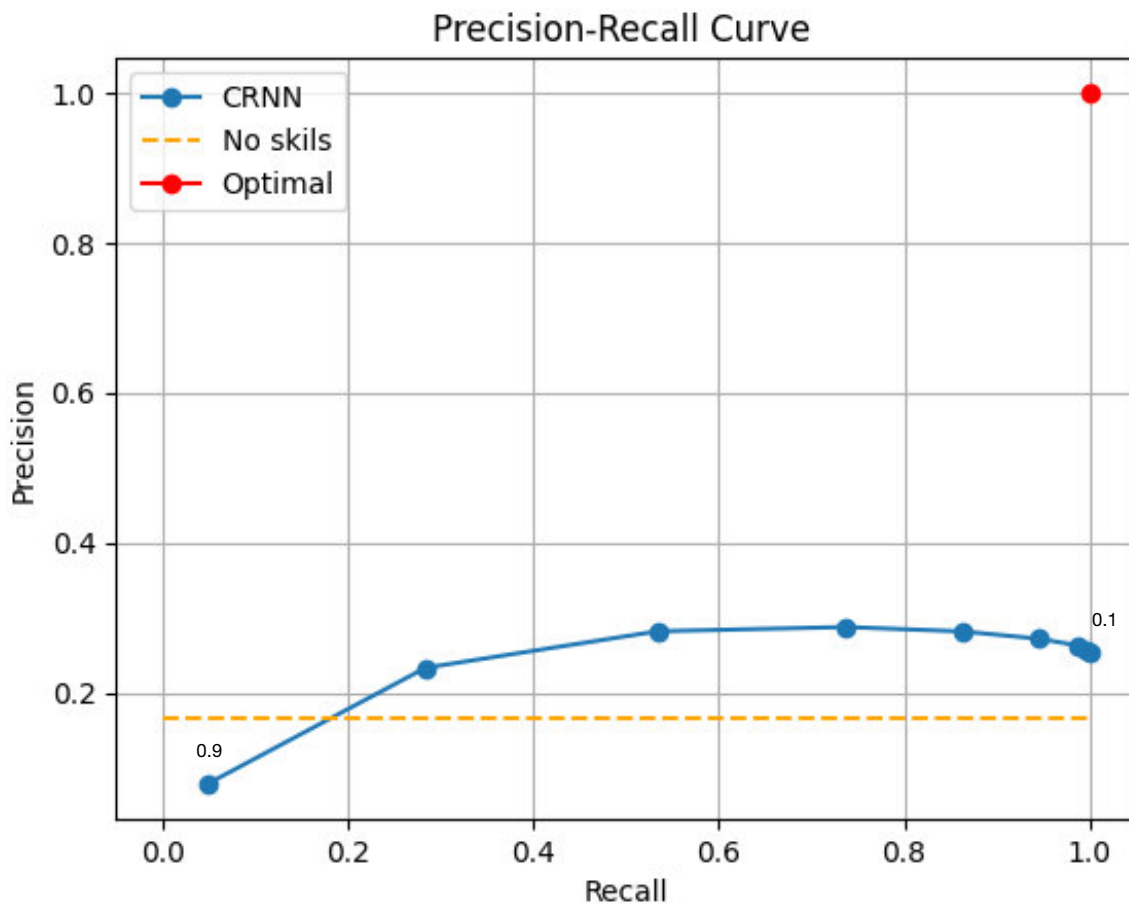


Figura 3.9: Esempio di curva precision-recall.

In conclusione, l'accuratezza, la recall e la precision sono metriche valutative essenziali per misurare le performance dei modelli di individuazione clinica. La scelta della metrica principale dipenderà dalle specifiche esigenze del contesto e dalla priorità assegnata alla riduzione dei falsi negativi rispetto ai falsi positivi. Nella maggior parte dei casi, la recall rappresenta la metrica chiave per garantire la massima sensibilità nella rilevazione delle condizioni cliniche, salvaguardando la sicurezza e la salute dei pazienti.

3.6 Validazione del modello

Nel perseguire l'obiettivo di ottenere risultati ottimali, è stato condotto un considerevole numero di esperimenti sia per la rilevazione delle apnee ostruttive che per la rilevazione delle apnee centrali. Di seguito verranno presentati solamente i risultati dei migliori esperimenti, selezionati in base alle performance più elevate durante la fase di validazione. Come parametri chiave per valutare l'efficacia del modello, abbiamo scelto di concentrarci principalmente sull'accuratezza e sulla recall. L'accuratezza è

una metrica molto informativa quando si lavora con dataset bilanciati, poiché misura il numero di predizioni corrette rispetto al totale delle predizioni effettuate. La recall è stata inclusa nella valutazione per la sua rilevanza nel contesto clinico come descritto nel paragrafo dedicato.

3.6.1 Apnee ostruttive

In relazione a questa specifica categoria di apnee, i risultati più significativi sono stati raggiunti in virtù dell'abbondante presenza di eventi di interesse all'interno della classe. Questo, per conseguenza, ha influito positivamente sulle prestazioni della rete, poiché ha comportato una maggior diversificazione dei dati in ingresso e, di conseguenza, una maggiore variabilità dei dati forniti alla rete neurale. Nelle figure 3.10 3.11 3.12 possiamo vedere le prestazioni raggiunte dal modello per la detection di apnee ostruttive del sonno, è da precisare inoltre che le linee in trasparenza che si vedono nei grafici rappresentano l'andamento reale delle curve mentre quelle a quelle in primo piano è stato applicato uno smoothing così da renderle meglio comprensibili.

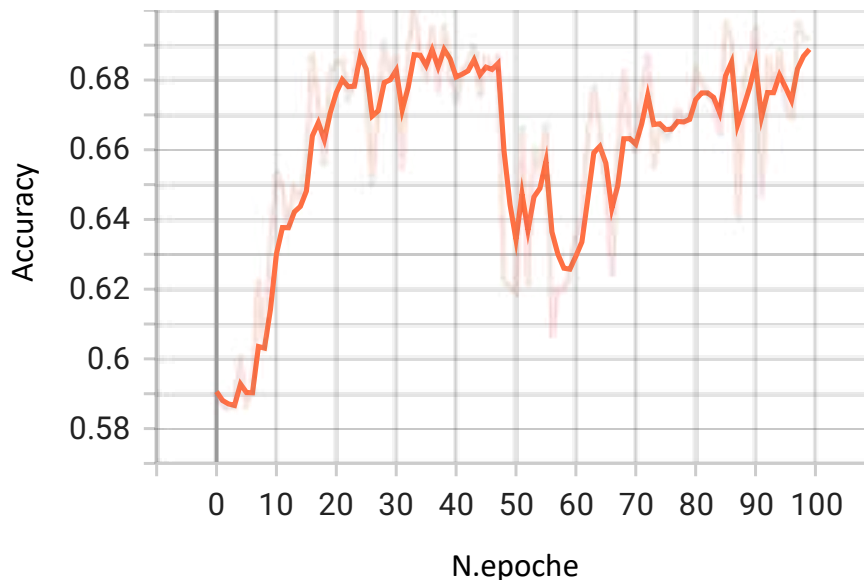


Figura 3.10: Accuratezza in fase di validazione

Come vediamo dai grafici il modello raggiunge un Accuratezza ed una Recall circa del 70% .

Certamente, è fondamentale riportare anche gli iperparametri scelti al fine dell'addestramento:

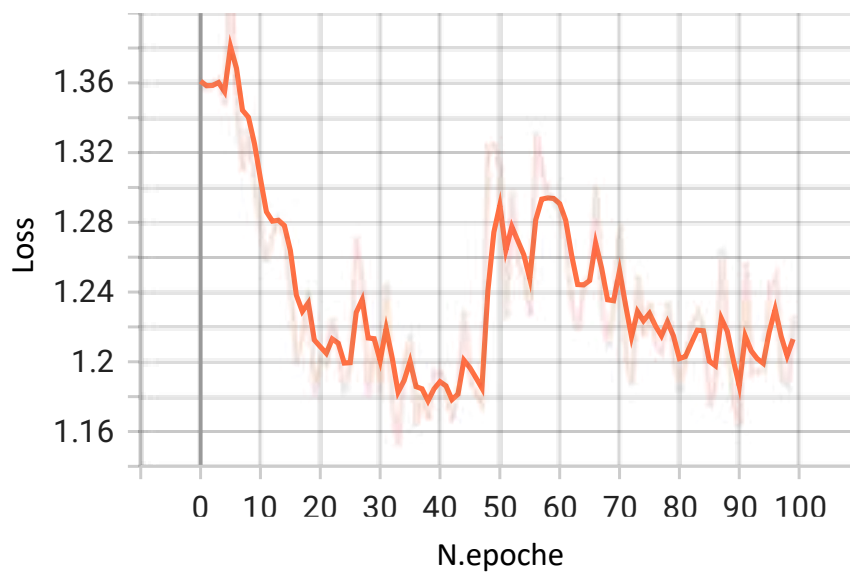


Figura 3.11: Loss in fase di validazione

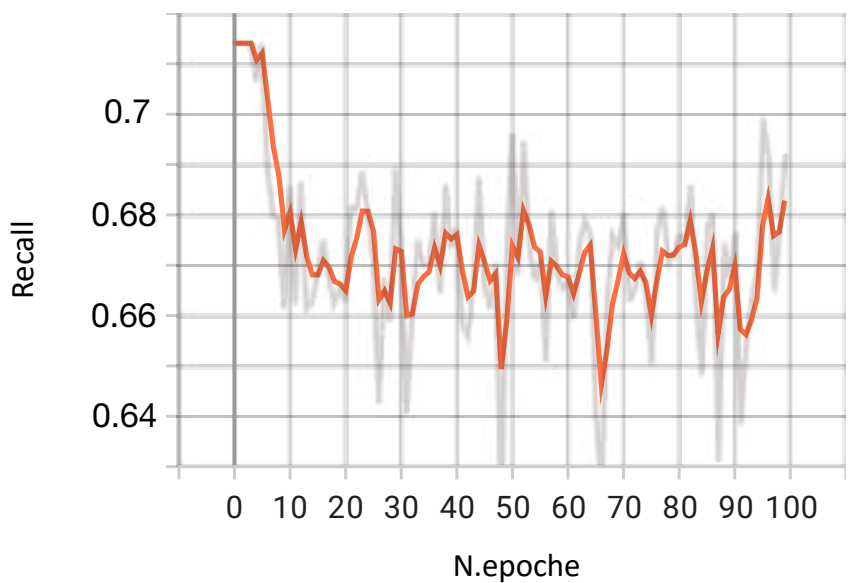


Figura 3.12: Recall in fase di validazione

- **Learning rate:** Il valore su cui lo scheduler lavora è compreso nel range (0.01-0.0001).
- **Batch size:** 8 in quanto con valori superiori si sperimentavano errori di allocazione di memoria.

- **Numero epoche:** 150 (limite per cui poi la rete entra in overfitting).
- **Win_length:** 50s.
- **thr:** 7s.
- **Apnea type:** ObstructiveApnea.

3.6.2 Apnee Centrali

Nel contesto delle apnee centrali, il modello ha conseguito risultati inferiori rispetto a quanto precedentemente delineato. Tale situazione si è verificata a causa della notevole scarsità degli eventi all'interno di questa categoria, in netto contrasto con le altre classi, e ciò ha sensibilmente limitato la variabilità intrinseca alla classe stessa. Di conseguenza, tali restrizioni nella diversità dei dati hanno esercitato un impatto sfavorevole sulle prestazioni globali della rete neurale.

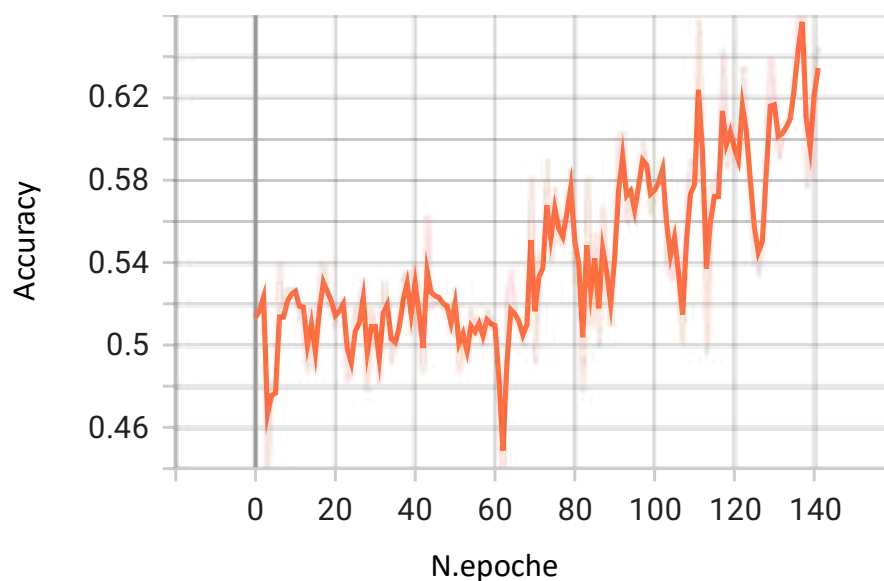


Figura 3.13: Accuratezza in fase di validazione

Come accennato in precedenza dai grafici 3.13 3.14 3.15 possiamo notare come accuratezza e recall siano nettamente più basse raggiungendo circa un 66% ciascuna.

Anche in questo caso si riportano gli iperparametri scelti al fine dell'addestramento del modello:

- **Learning rate:** Il valore su cui lo scheduler lavora è compreso nel range (0.01-0.0001).

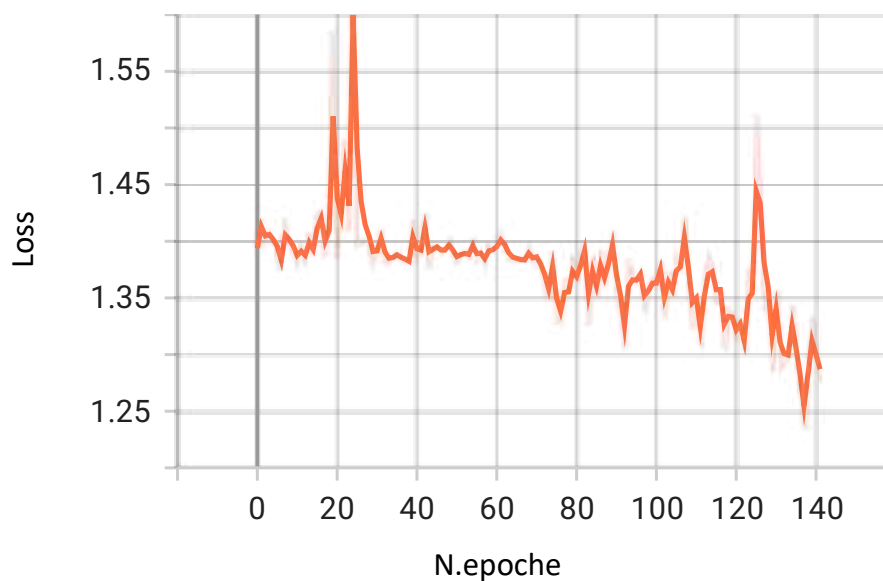


Figura 3.14: Loss in fase di validazione

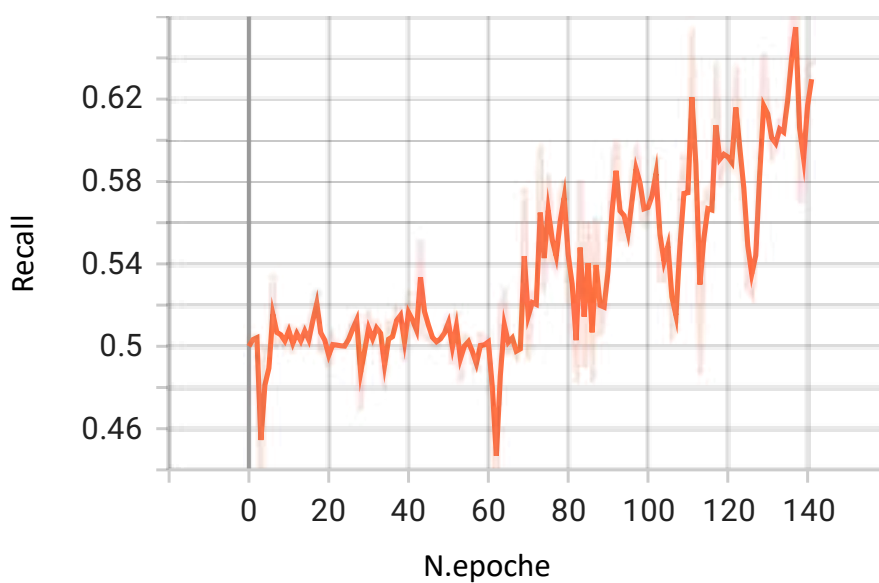


Figura 3.15: Loss in fase di validazione

- **Batch size:** 8 in quanto con valori superiori si sperimentavano errori di allocazione di memoria.
- **Numero epoche:** 150 (limite per cui poi la rete entra in overfitting).

- **Win_lengh:**50s.
- **thr:** 7s.
- **Apnea type:** CentralApnea.

3.6.3 Preparazione del test set

Durante la fase di testing del modello abbiamo adottato un approccio che mirava a simulare il più fedelmente possibile una situazione clinica reale. In termini pratici, ciò significa che abbiamo effettuato il test del modello non selezionando casualmente campioni di dati relativi all'apnea e non apnea dal dataset, ma valutando in modo completo e sequenziale l'intera registrazione dei cicli del sonno.

Gli spettrogrammi estratti da queste registrazioni sono stati inviati singolarmente alla rete, senza essere raggruppati in batch. Inoltre, è importante notare che gli spettrogrammi sono stati generati con un overlap del 96%, il che implica che per ogni file di 5 minuti (300 secondi), sono stati generati 125 spettrogrammi. Questo ampio overlap ci ha permesso di effettuare predizioni basate su una finestra temporale di appena due secondi tra uno spettrogramma e l'altro aumentando molto la sensibilità della rete. Tuttavia, la valutazione di finestre da 50s consente comunque di effettuare predizioni basandosi su finestre temporali ampie, permettendoci di contestualizzare l'apnea quando è presente nei dati.

Per migliorare le performance del classificatore in post-processing si è deciso applicare un filtraggio così da correggere buona parte delle predizioni che risultano essere fuori contesto rispetto alle predizioni precedenti o successive.

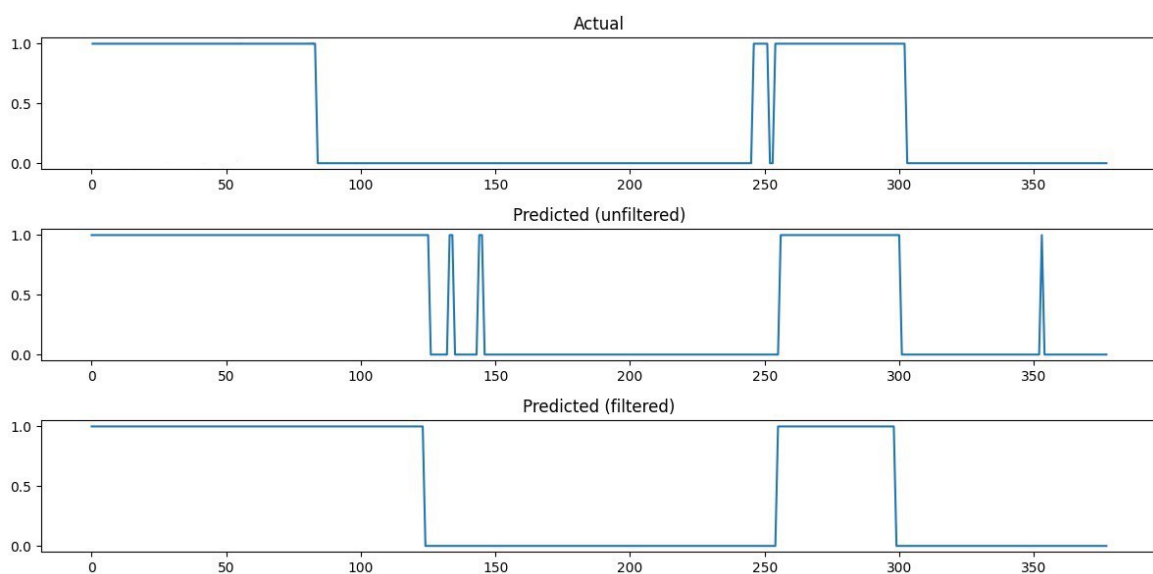


Figura 3.16: Filtraggio falsi positivi

Come mostrato nell'immagine 3.16, nel primo riquadro vengono presentati i valori effettivi estratti dalle label, contrassegnati come "Actual". Subito sotto, sono riportate le predizioni fornite dalla rete neurale. È evidente dalla figura che la rete ha prodotto una serie di falsi positivi molto localizzati, i quali hanno avuto un impatto negativo significativo sulle prestazioni complessive del sistema di classificazione.

Di conseguenza, è stato ritenuto necessario intraprendere misure correttive al fine di eliminare queste fluttuazioni anomale, che possono risultare fuorvianti nella fase di classificazione. In tal senso, è stato applicato un filtro passa-basso a media mobile per eliminare le predizioni anomale (33). I risultati di questo processo sono presentati nell'ultimo grafico dell'immagine, da cui emerge che tutti i falsi positivi sono stati correttamente eliminati.

Tuttavia, è importante notare che in caso di falsi positivi di lunga durata, il filtro può contribuire a ridurre l'errore, ma non sarà in grado di eliminarlo completamente. Questo filtro ha dimostrato di migliorare in modo significativo le prestazioni complessive della rete, ma non può risolvere completamente il problema in caso di falsi positivi prolungati.

I test sono stati eseguiti sia su base individuale, considerando ciascun paziente separatamente, che sull'intero dataset di test.

Inoltre, è stata condotta un'analisi statistica per determinare la migliore soglia di decisione del classificatore. A questo scopo, sono state plottate le curve Precision-Recall al variare della soglia, consentendo di identificare la soglia ottimale che massimizza contemporaneamente sia la Recall che la Precision. La valutazione della soglia ottimale è stata effettuata calcolando la distanza minima tra il punto di ottimo e quello identificato dalla soglia. Una discussione dettagliata di queste curve sarà fornita in seguito, nel capitolo dedicato alla discussione dei risultati.

Capitolo 4

Risultati e discussione

In questo capitolo, saranno esposti i risultati conseguiti dal modello, precedentemente addestrato, nel riconoscimento delle due tipologie di apnea, ossia quella centrale e quella ostruttiva. In seguito, saranno esaminati gli effetti delle tecniche di post-processing impiegate, seguiti da un'analisi qualitativa volta a comprendere le ragioni per cui alcuni pazienti ottengono prestazioni inferiori alla media durante la fase di test. Infine, verrà condotto un confronto tra i risultati ottenuti nel presente studio e quelli presentati nell'articolo di riferimento (17), che fa uso dello stesso dataset.

4.1 Apnee ostruttive

Nel paragrafo presente, presenteremo e discuteremo i risultati ottenuti dal modello nella classificazione delle apnee ostruttive del sonno. Come già evidenziato nel paragrafo dedicato, durante la fase di addestramento il modello ha raggiunto una recall del 70,5%, un'accuratezza del 70% e una precision del 53,3%, mentre nella fase di convalida abbiamo ottenuto una recall del 69%, una precision del 47% e un'accuratezza del 70%.

È importante sottolineare che, in questa fase, abbiamo un dataset forzato ad essere bilanciato per cui l'accuratezza ricopre un ruolo fondamentale nella valutazione del modello, il che non accade durante la fase di test, poiché il dataset lì risulta essere pesantemente sbilanciato.

Le prestazioni non ottimali del modello addestrato possono essere attribuite principalmente al fatto che il dataset in uso è molto rumoroso e contiene una notevole quantità di dati altamente corrotti (figura 4.3). Questi dati compromettono negativamente il processo di addestramento del modello.

Nel tentativo di migliorare la qualità dei dati, è stata esaminata la possibilità di eliminare o ridurre il rumore di fondo, ma questo approccio ha prodotto risultati scarsi. Il motivo è che il filtraggio del rumore risulta dannoso anche per il segnale

utile, compromettendo così la capacità del modello di identificare correttamente le apnee ostruttive del sonno.

È importante sottolineare che la gestione del rumore nei dati medici può essere una sfida complessa, poiché è necessario trovare un equilibrio tra la rimozione del rumore e la conservazione delle informazioni cliniche rilevanti. Questo potrebbe richiedere ulteriori sforzi nella pulizia e nella preparazione dei dati o l'uso di tecniche più avanzate di elaborazione del segnale per migliorare le prestazioni del modello. Tuttavia, dal momento che i task di riduzione del rumore sono molto complessi, si è deciso di ometterli dalla trattazione.

Per quanto riguarda la fase di test, di seguito verranno presentate le prestazioni del modello dopo il post-processing dei dati in uscita dalla rete. Dato che il dataset è fortemente sbilanciato in questa fase, l'accuratezza risulta essere non informativa nella valutazione del modello. Si preferisce invece utilizzare le curve Precision-Recall, poiché rappresentano in modo più informativo le prestazioni ottenute.

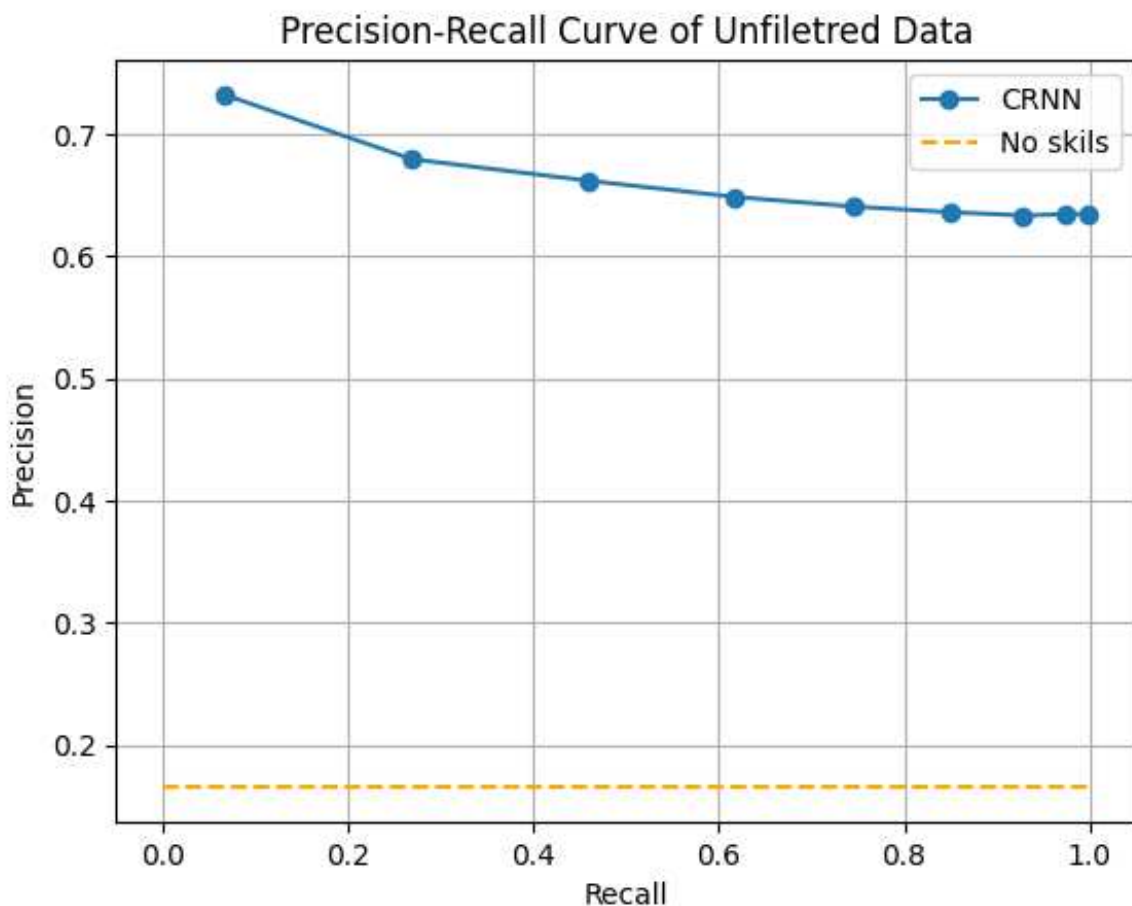


Figura 4.1: Curva precision-recall

Analizzando il grafico presente nella Figura 4.1, è evidente che le prestazioni del nostro modello sono notevolmente superiori rispetto alla linea di "no-skill" (nessuna

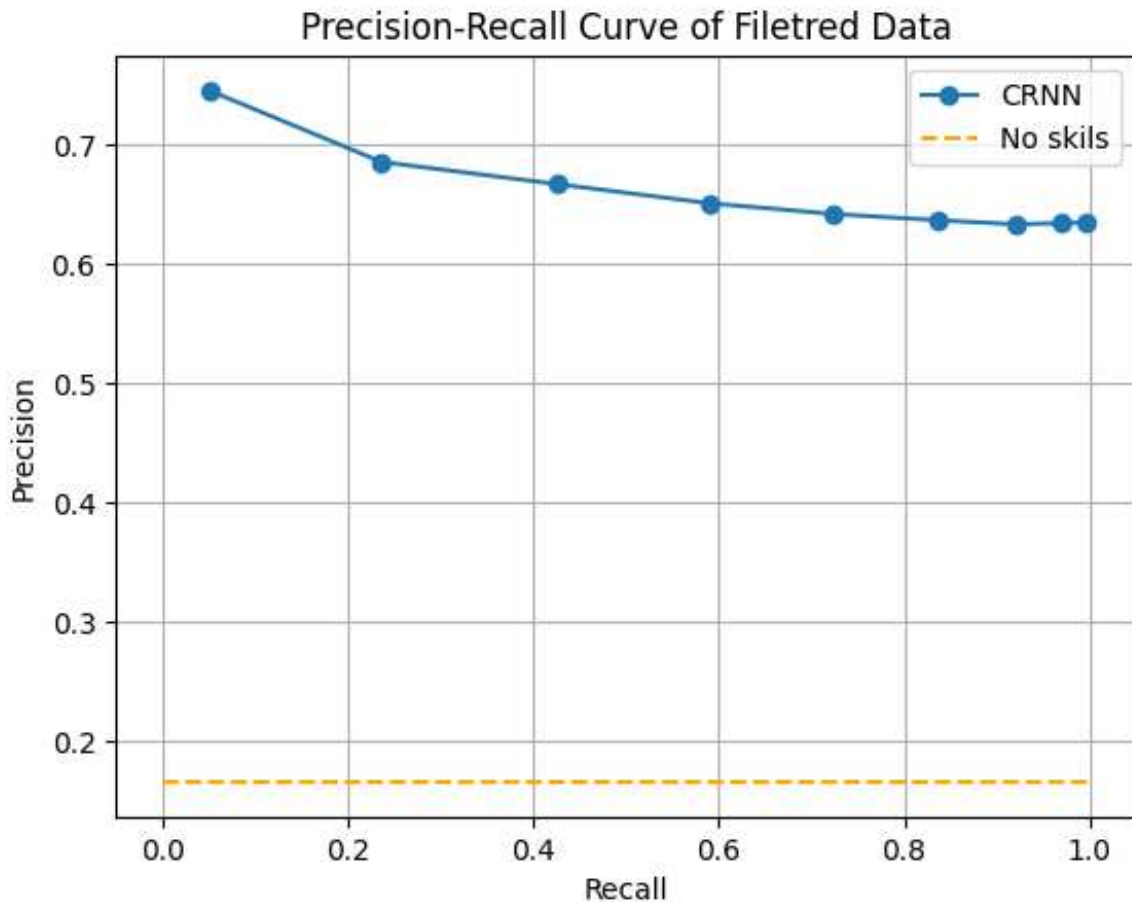


Figura 4.2: Curva precision-recall dopo il post-processing

capacità) posta ad un valore di 0.166, questa rappresenta i valori per i quali il classificatore opera come un "random guess" (indovino casuale). Questa soglia viene calcolata come:

$$\text{No Skils} = \frac{\text{numero elementi classe apnea}}{\text{numero totale di elementi nel dataset}} \quad (4.1)$$

Dal grafico si evince anche che esiste un trade-off tra recall e precision cioè massimizzando la recall la precision viene meno e viceversa se si tenta di massimare la precision. Dall'analisi risultante la soglia migliore è quella con la minor distanza dal punto di ottimo cioè 0.1: così facendo si avrà un numero elevato di falsi positivi. In ambito clinico comunque ciò non rappresenta una limitazione, bensì consente di non sotto-stimare gli eventi di apnea ed evitare possibili conseguenze negative per il paziente. Fatto riferimento alla Figura 4.2, è possibile osservare che l'applicazione di un filtro in post-processing contribuisce a un modesto miglioramento dei risultati in termini di precision che raggiunge il 64%, mantenendo invariata la recall che rimane al 99,4%. Tale effetto si configura come desiderato. Infine è importante fare una digressione

riguardante la valutazione del modello per singoli pazienti, in questa casistica risulta essere comune in fase di test trovare paziente che restituiscono metriche molto al di sotto delle medie mostrate sopra. Da un'analisi qualitativa delle possibili cause di questo fenomeno è risultato evidente che una possibile causa sia dovuta al rumore presente negli spettrogrammi che risulta essere più accentuato per alcuni pazienti. Un esempio viene mostrato in figura 4.3 mentre in 4.4 viene mostrato un esempio di spettrogramma con una presenza di rumore molto più bassa¹.

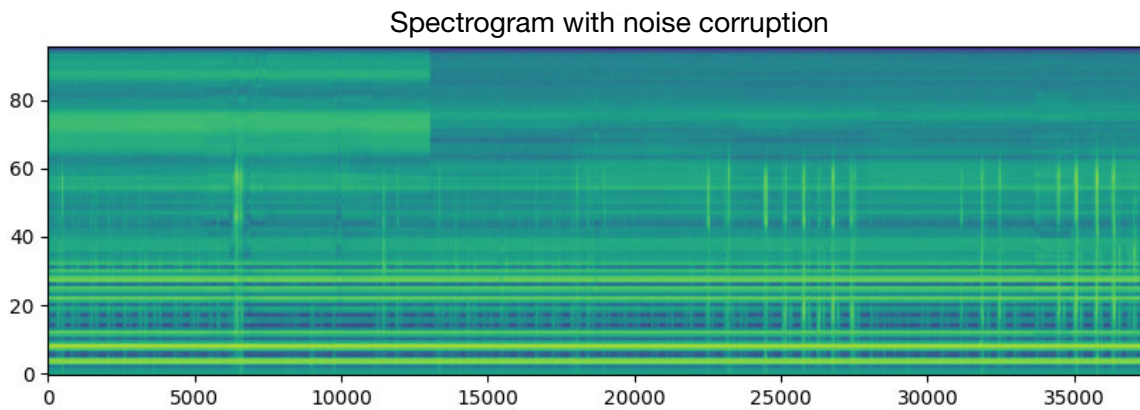


Figura 4.3: Esempio di spettrogramma con presenza di rumore invalidante

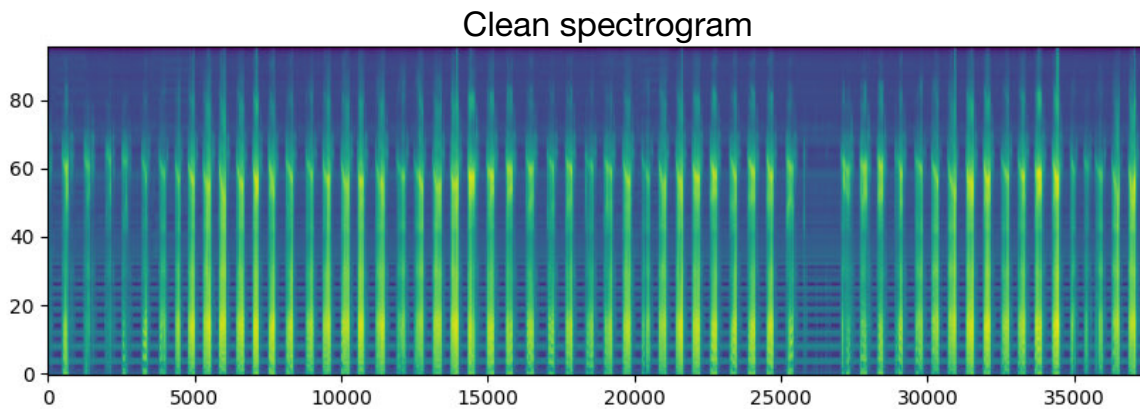


Figura 4.4: Esempio di spettrogramma con una presenza di rumore non invalidante

Quando confrontiamo le prestazioni del nostro modello con l'articolo di riferimento di Korompili et al. (2021) (17), notiamo che il nostro modello ottiene notevoli miglioramenti in termini di prestazioni. Tuttavia, è fondamentale considerare alcune importanti considerazioni:

- **Disponibilità dei dati:** Gli autori dell'articolo di riferimento non condividono il dataset completo e non specificano in dettaglio come è stato costituito il dataset

¹Importare sottolineare il fatto che la generazione degli spettrogrammi è avvenuta con la stessa normalizzazione così da poter renderli confrontabili.

di test. Questo rende difficile la comparazione diretta poiché non si dispone di tutti i dettagli necessari sulla qualità e la rappresentatività del loro dataset.

- **Approccio diverso:** Il nostro modello utilizza un approccio differente rispetto a quello presentato dagli autori dell'articolo. Nell'articolo (17) viene proposto un approccio probabilistico basato su un voice activity detector (VAD) e successivamente applicano un sistema di soglie in post-processing.
- **Metriche valutative limitate:** Nell'articolo viene riportata esclusivamente la recall come parametro di valutazione.
- **Valutazione critica:** È essenziale effettuare il confronto con una certa cautela, tenendo conto delle differenze nell'approccio e nella disponibilità dei dati. I risultati devono essere interpretati alla luce di queste differenze, e l'effettiva validità della comparazione potrebbe richiedere ulteriori dettagli e analisi.

Nell'articolo Korompili et. al. riportano un valore di Recall del 77.9% per le apnee ostruttive, mentre il nostro modello raggiunge un valore pari al 99.4%.

4.2 Apnee Centrali

In questa istanza, in confronto alla precedente, si osserva un decremento delle prestazioni nel processo di addestramento, con un'accuratezza e una Recall che raggiungono rispettivamente il 68% e il 64% in fase di validazione. Questa diminuzione delle prestazioni può essere attribuita alla limitata disponibilità di dati nella classe "CentralApnea", che fornisce un numero ridotto di campioni per l'addestramento del modello, rendendo notevolmente complesso il compito di addestrare un modello più performante.

Le curve precision-recall sono ora presentate per l'output del modello prima dell'applicazione del filtraggio (Figura 4.5) e dopo il filtraggio (Figura 4.6). Si può confermare che le stesse considerazioni fatte per le curve relative alle apnee ostruttive mantengono la loro validità anche in questo contesto.

Infatti in questa casistica, si osserva che la curva precision-recall è significativamente superiore al limite del "no skill" (0,073). Tuttavia, concentrandoci sulla curva ottenuta dopo l'applicazione del post-processing dei dati in questo specifico caso, non si notano miglioramenti sostanziali. Ciò è dovuto al fatto che, rispetto al modello per le apnee ostruttive, il modello in esame presenta prestazioni inferiori durante il processo di addestramento, il che comporta un notevole aumento di falsi positivi/negativi che riducono l'efficacia del filtro. Anche in questo caso la soglia ottimale risulta essere nella parte destra del grafico corrispondente alla soglia di decisione pari a 0.1 che

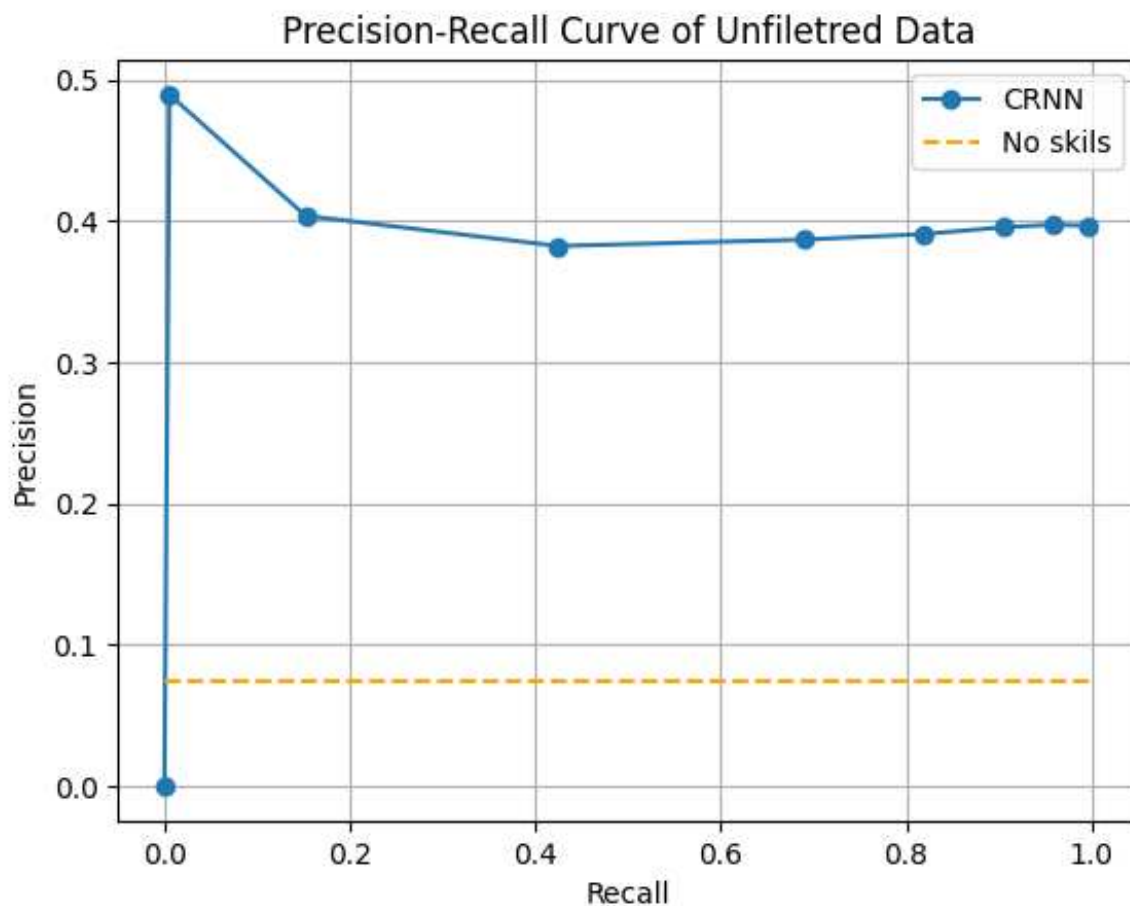


Figura 4.5: Curva precision-recall per apnee centrali

raggiunge valori di precision e recall rispettivamente pari al 39,8% e al 99,1%. Inoltre, durante l'analisi paziente per paziente, sono state riscontrate incongruenze tra i risultati e le metriche generali calcolate sul dataset di test. Questo fenomeno può ancora essere attribuito alla presenza di un rumore di fondo significativo all'interno delle registrazioni.

Volendo confrontare le prestazioni ottenute da Korompili et.al. nella detection delle apnee centrali con i nostri risultati osserviamo che nell'articolo di Korompili et.al. si raggiunge un valore di recall dell'88,3% contro il nostro 99,1%.

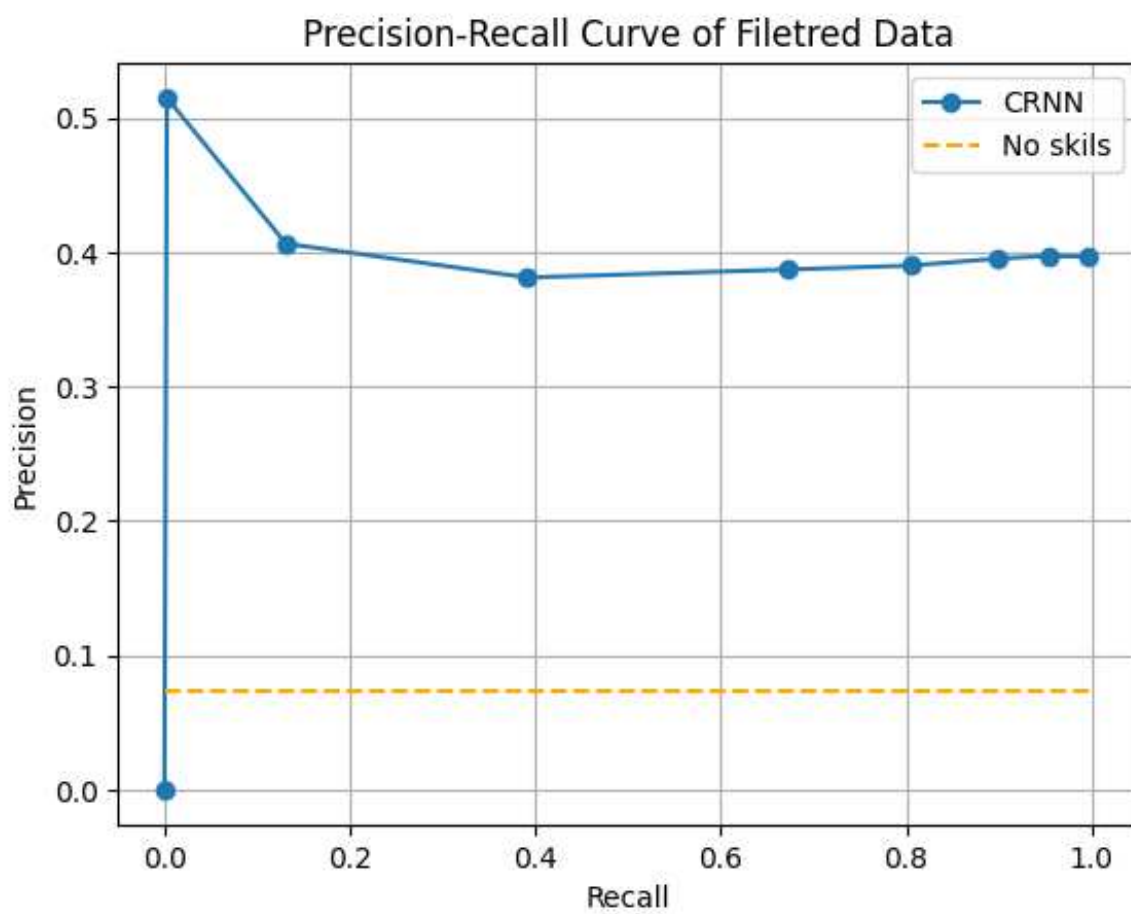


Figura 4.6: Curva precision-recall per apnee centrali dopo il post-processing

Capitolo 5

Conclusioni e sviluppi futuri

Nei capitoli della tesi, sono state introdotte ed esaminate alcune questioni relative all'identificazione delle apnee notturne, concentrandosi in particolare sulle tipologie più problematiche dal punto di vista clinico, ossia le apnee ostruttive e Centrali. Dopo una dettagliata esposizione della sindrome per entrambe le categorie di interesse, sono state approfondite l'epidemiologia, la descrizione clinica degli eventi, i fattori di rischio e le metodologie di identificazione.

Nel secondo capitolo, è stata fornita una dettagliata panoramica dello stato dell'arte, distinguendo tra i metodi tradizionali e le tecniche basate sul deep learning utilizzate per supportare l'identificazione di tali eventi. Successivamente, sono stati delineati gli obiettivi specifici di questa tesi.

Il terzo capitolo è stato dedicato allo studio dei materiali e dei metodi impiegati. Inizialmente, è stato presentato il dataset nella sua completezza, con dettagli sulle attività di pre-processing utilizzate e un'analisi statistica per visualizzare la distribuzione e la durata degli eventi. Successivamente, è stata descritta la creazione del dataset di addestramento, le metriche di valutazione utilizzate, la validazione del modello e la strutturazione del dataset di test, comprese le operazioni di post-processing.

Nel quarto capitolo, sono stati illustrati e discussi i risultati ottenuti, distinguendo tra i due tipi di apnee. È stato infatti evidenziato come la nostra metodologia risulta più performante rispetto a quella proposta da Korompili et.al.(17) raggiungendo il 99,4% di recall e il 64% di precision per le SAS ostruttive; mentre nelle CSA si arriva ad un 99,1% di recall ed un 39,8% di precision.

I risultati ottenuti consentono di concludere che l'attuale tecnologia basata sul deep learning non è ancora sufficientemente matura per garantire una rilevazione precisa e affidabile delle apnee notturne. Tuttavia, essa si presenta come uno strumento valido per supportare il clinico nell'accelerare il processo di identificazione.

I possibili sviluppi futuri si concentreranno sull'obiettivo di ottenere un dataset più pulito, riducendo al minimo il rumore di fondo, e sull'adozione di architetture di rete

più avanzate e profonde al fine di migliorare il processo di estrazione delle caratteristiche, che in un compito così complesso riveste un ruolo fondamentale. Inoltre, verranno esplorate altre strategie per la creazione del dataset utilizzato per l'addestramento della rete e verranno proposti algoritmi di post-processing per affinare i dati prodotti in uscita dalla rete.

In definitiva, questa ricerca contribuisce allo sviluppo tecnologico dietro l'identificazione delle apnee, offrendo importanti spunti per il miglioramento delle tecnologie e delle metodologie utilizzate in campo clinico. La lotta contro le apnee notturne rimane una sfida importante per la medicina, e le ricerche future potrebbero portare a progressi significativi nella diagnosi e nel trattamento di questa condizione medica.

Bibliografia

- [1] J. V. Rundo and R. Downey III, “Polysomnography,” *Handbook of clinical neurology*, vol. 160, pp. 381–392, 2019.
- [2] D. G. Ingram, S. C. Crane, and A. C. Halbower, “Polysomnography,” *Sleep in Children with Neurodevelopmental Disabilities: An Evidence-Based Guide*, pp. 27–43, 2019.
- [3] G. Korompili, A. Amfilochiou, L. Kokkalas, S. A. Mitilneos, N.-A. Tatlas, M. Kouvaras, E. Kastanakis, C. Maniou, and S. M. Potirakis, “Psg-audio, a scored polysomnography dataset with simultaneous audio recordings for sleep apnea studies,” *Scientific data*, vol. 8, no. 1, p. 197, 2021.
- [4] H. B. Parvaneh Janbakhshi, Ina Kodrasi, “Obstructive sleep apnea syndrome (osas). review of the literature,” pp. 6405–6409, 2019.
- [5] L. RC., “Obstructive sleep apnea syndrome (osas) in children. diagnostic challenges.” 1996.
- [6] S. M. Caples, A. S. Gami, and V. K. Somers, “Obstructive sleep apnea,” *Annals of internal medicine*, vol. 142, no. 3, pp. 187–197, 2005.
- [7] E. A. Bresnitz, R. Goldberg, and R. M. Kosinski, “Epidemiology of obstructive sleep apnea.” *Epidemiologic reviews*, vol. 16, no. 2, pp. 210–227, 1994.
- [8] N. M. Punjabi, “The epidemiology of adult obstructive sleep apnea,” *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 136–143, 2008.
- [9] T. Young, P. E. Peppard, and D. J. Gottlieb, “Epidemiology of obstructive sleep apnea: a population health perspective,” *American journal of respiratory and critical care medicine*, vol. 165, no. 9, pp. 1217–1239, 2002.
- [10] P. M. H. K. F. L. M. B. e. a. Young T, Peppard P, “Population-based study of sleep-disordered breathing as a risk factor for hypertension. 1997;157:1746-52.” *Arch Intern Med*, 1997.

- [11] D. J. Eckert, A. S. Jordan, P. Merchia, and A. Malhotra, “Central sleep apnea: pathophysiology and treatment,” *Chest*, vol. 131, no. 2, pp. 595–607, 2007.
- [12] S. Javaheri, “Central sleep apnea,” *Clinics in chest medicine*, vol. 31, no. 2, pp. 235–248, 2010.
- [13] S. Zeineddine and M. S. Badr, “Treatment-emergent central apnea: physiologic mechanisms informing clinical practice,” *Chest*, vol. 159, no. 6, pp. 2449–2457, 2021.
- [14] A. Salloum, J. A. Rowley, J. H. Mateika, S. Chowdhuri, Q. Omran, and M. S. Badr, “Increased propensity for central apnea in patients with obstructive sleep apnea: effect of nasal continuous positive airway pressure,” *American journal of respiratory and critical care medicine*, vol. 181, no. 2, pp. 189–193, 2010.
- [15] B. Jafari and V. Mohsenin, “Polysomnography,” *Clinics in chest medicine*, vol. 31, no. 2, pp. 287–297, 2010.
- [16] C. Baglioni, W. Regen, A. Teghen, K. Spiegelhalder, B. Feige, C. Nissen, and D. Riemann, “Sleep changes in the disorder of insomnia: a meta-analysis of polysomnographic studies,” *Sleep medicine reviews*, vol. 18, no. 3, pp. 195–213, 2014.
- [17] G. Korompili, L. Kokkalas, S. A. Mitilneos, N.-A. Tatlas, and S. M. Potirakis, “Detecting apnea/hypopnea events time location from sound recordings for patients with severe or moderate sleep apnea syndrome,” *Applied Sciences*, vol. 11, no. 15, p. 6888, 2021.
- [18] B. Wang, X. Tang, H. Ai, Y. Li, W. Xu, X. Wang, and D. Han, “Obstructive sleep apnea detection based on sleep sounds via deep learning,” *Nature and Science of Sleep*, pp. 2033–2045, 2022.
- [19] H. Nakano, T. Furukawa, and T. Tanigawa, “Tracheal sound analysis using a deep neural network to detect sleep apnea,” *Journal of Clinical Sleep Medicine*, vol. 15, no. 8, pp. 1125–1133, 2019.
- [20] R. Li, W. Li, K. Yue, and Y. Li, “Convolutional neural network for screening of obstructive sleep apnea using snoring sounds,” *Biomedical Signal Processing and Control*, vol. 86, p. 104966, 2023.
- [21] S. Cheng, C. Wang, K. Yue, R. Li, F. Shen, W. Shuai, W. Li, and L. Dai, “Automated sleep apnea detection in snoring signal using long short-term memory

- neural networks,” *Biomedical Signal Processing and Control*, vol. 71, p. 103238, 2022.
- [22] E. Cakır, G. Parascandolo, T. Heittola, H. Huttunen, and T. Virtanen, “Convolutional recurrent neural networks for polyphonic sound event detection,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 6, pp. 1291–1303, 2017.
- [23] F. Grondin, J. Glass, I. Sobieraj, and M. D. Plumbley, “Sound event localization and detection using crnn on pairs of microphones,” *arXiv preprint arXiv:1910.10049*, 2019.
- [24] D. Basaran, S. Essid, and G. Peeters, “Main melody extraction with source-filter nmf and crnn,” in *19th International Society for Music Information Retrieval*, 2018.
- [25] J. Sang, S. Park, and J. Lee, “Convolutional recurrent neural networks for urban sound classification using raw waveforms,” in *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE, 2018, pp. 2444–2448.
- [26] F. M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed, “A comprehensive overview and comparative analysis on deep learning models: Cnn, rnn, lstm, gru,” *arXiv preprint arXiv:2305.17473*, 2023.
- [27] R. Lu and Z. Duan, “Bidirectional gru for sound event detection,” *Detection and Classification of Acoustic Scenes and Events*, pp. 1–3, 2017.
- [28] M. Hossin and M. N. Sulaiman, “A review on evaluation metrics for data classification evaluations,” *International journal of data mining & knowledge management process*, vol. 5, no. 2, p. 1, 2015.
- [29] J. Miao and W. Zhu, “Precision–recall curve (prc) classification trees,” *Evolutionary intelligence*, vol. 15, no. 3, pp. 1545–1569, 2022.
- [30] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich, “The area under the precision–recall curve as a performance metric for rare binary events,” *Methods in Ecology and Evolution*, vol. 10, no. 4, pp. 565–577, 2019.
- [31] M. Johansson, “The hilbert transform,” *Mathematics Master’s Thesis. Växjö University, Suecia. Disponible en internet: http://w3.msi.vxu.se/exarb/mj_ex.pdf, consultado el*, vol. 19, 1999.
- [32] F. R. Kschischang, “The hilbert transform,” *University of Toronto*, vol. 83, p. 277, 2006.

- [33] B. Justusson, “Median filtering: Statistical properties,” *Two-Dimensional Digital Signal Processing II: Transforms and Median Filters*, pp. 161–196, 2006.
- [34] H. H. Tan and K. H. Lim, “Vanishing gradient mitigation with deep learning neural network optimization,” in *2019 7th international conference on smart computing & communications (ICSCC)*. IEEE, 2019, pp. 1–4.
- [35] M. El Shayeb, L.-A. Topfer, T. Stafinski, L. Pawluk, and D. Menon, “Diagnostic accuracy of level 3 portable sleep tests versus level 1 polysomnography for sleep-disordered breathing: a systematic review and meta-analysis,” *Cmaj*, vol. 186, no. 1, pp. E25–E51, 2014.
- [36] S.-J. Park, K.-H. Park, D.-J. Shin, H.-M. Park, Y.-B. Lee, and Y.-H. Sung, “Comparing sleep questionnaires and respiratory disturbance index in obstructive sleep apnea syndrome patients,” *Journal of Korean Sleep Research Society*, vol. 5, no. 2, pp. 89–93, 2008.