

UNIVERSITÀ DI PAVIA

FACOLTÀ DI INGEGNERIA

DIPARTIMENTO DI INGEGNERIA INDUSTRIALE E
DELL'INFORMAZIONE

CORSO DI LAUREA MAGISTRALE IN
BIOINGEGNERIA

TESI DI LAUREA

**Analisi della variabilità della frequenza
cardiaca mediante Anomaly Detection per
predire stati aggressivi in soggetti autistici**

Relatore:

Prof. Pietro Savazzi

Correlatore:

Prof.ssa Natascia Brondino

Candidata:

Alessia Tudino

A.A. 2024/2025

Indice

Elenco degli Acronimi	9
Abstract	13
1 Introduzione	15
1.1 Il problema	15
1.2 I precursori	16
1.3 L'obiettivo e il contesto di ricerca	17
1.4 Struttura dell'elaborato	18
2 Stato dell'arte	21
2.1 Il SNA come base fisiologica della regolazione emotiva nell'ASD . . .	21
2.1.1 I domini di analisi dell'HRV	21
2.1.2 La teoria polivagale e il modello di Beauchaine	24
2.2 Evidenze empiriche sulla risposta cardiovascolare ai fattori di stress nell'ASD	27
2.3 Dall'epidemiologia alla fisiologia: i fattori di rischio per l'aggressività nell'ASD	29
2.4 Dalla misurazione alla predizione: dispositivi indossabili e algoritmi automatici	31
2.5 La Change Point Analysis come strumento per rilevare i precursori . . .	33
2.6 Basi fisiologiche dello stress operativo nei piloti di aeromobili	35
2.7 Rilevamento del mental workload nei piloti tramite segnali fisiologici	38
2.8 Posizionamento del presente lavoro nella letteratura	40
3 Dataset e acquisizione	43
3.1 Il campione di studio e il contesto sperimentale	43
3.2 Strumentazione di acquisizione	44
3.3 Protocollo e sessioni sperimentali	47
3.3.1 Acquisizioni del 16 novembre 2022	48
3.3.2 Acquisizioni del 18 novembre 2022	48

3.3.3	Dal protocollo cartaceo al protocollo digitale	49
3.3.4	Riepilogo delle sessioni analizzate	51
3.4	Criticità nell'acquisizione dei dati	52
4	Pre-processing	55
4.1	Caricamento e segmentazione	55
4.2	Decomposizione dell'EDA	57
4.3	Elaborazione dell'accelerometro	58
4.4	Preparazione del segnale HRV	58
4.5	Qualità del dato finale	59
4.6	Analisi qualitativa dei segnali pre-processati	60
4.6.1	Sessioni del 16 novembre 2022	60
4.6.2	Sessioni del 18 novembre 2022	64
5	Estrazione delle feature	69
5.1	Calibrazione della baseline individuale	69
5.2	Segmentazione a finestre scorrevoli	69
5.3	Feature estratte	70
5.3.1	Feature di frequenza cardiaca	70
5.3.2	Feature EDA tonica	71
5.3.3	Feature EDA fasica e driver	72
5.3.4	Feature di movimento	74
5.3.5	Variabilità della frequenza cardiaca	74
5.4	Dataset finale	75
5.5	Distribuzione statistica delle feature nel dataset	76
5.5.1	Analisi per feature di frequenza cardiaca	76
5.5.2	Analisi per feature EDA tonica	76
5.5.3	Analisi per feature EDA fasica e driver	77
5.5.4	Analisi per feature ACC e HRV	77
5.5.5	Considerazioni sulla separabilità del dataset	77

6	Modelli predittivi	79
6.1	Modello non supervisionato: Anomaly Detection	79
6.1.1	Motivazione della scelta	79
6.1.2	Architettura dell'algoritmo	80
6.1.3	Risultati e validazione clinica	81
6.1.4	Modello supervisionato: Random Forest soggetto-specifico	90
7	Modello ibrido per la predizione di crisi	95
7.1	Architettura del modello ibrido	95
7.2	Generazione dei dati sintetici	96
7.2.1	Struttura della sessione simulata	96
7.3	Anomaly detection e Change Point Analysis	97
7.3.1	Pre-elaborazione del segnale EDA	97
7.3.2	Rilevamento del change point	98
7.3.3	Validazione clinica multi-segnale	98
7.4	Classificatore Random Forest	99
7.4.1	Definizione delle etichette	99
7.4.2	Addestramento	99
7.4.3	Predizione e soglia di allarme	99
7.5	Risultati del modello ibrido	100
7.5.1	Dati artificiali, change point e predizione	100
7.5.2	Matrice di confusione	101
7.5.3	Curva ROC e AUC	101
7.5.4	Importanza delle feature	102
7.5.5	Confronto training vs test	103
7.6	Validazione della robustezza temporale	104
7.6.1	Robustezza della Change Point Analysis	104
7.6.2	Robustezza del Random Forest	106
7.7	Analisi di robustezza al rumore	107
7.7.1	Il rumore nei segnali fisiologici	107
7.7.2	Il test di stress a +6 dB	108

7.7.3	Risultati del test di stress	108
7.8	Soluzione al degrado di sensibilità	113
7.8.1	Interventi implementati	113
7.8.2	Risultati del modello robusto	114
7.9	Prospettive di trasferibilità clinica: dai dati sintetici al contesto reale .	119
7.9.1	Pre-elaborazione avanzata: scomposizione dell'EDA e filtraggio degli artefatti	119
7.9.2	Dalle soglie globali alle finestre temporali mobili	120
7.9.3	Standardizzazione intra-soggetto per il Random Forest	120
7.9.4	Transizione dall'analisi offline all'elaborazione in streaming .	121
7.9.5	Transfer learning: integrare la simulazione con i dati clinici reali	123
8	Generalizzazione del modello ibrido ai piloti aerei	125
8.1	Basi fisiologiche: ASD vs piloti	125
8.2	Giustificazione dell'approccio sintetico	126
8.3	Generazione dei dati sintetici	127
8.4	Architettura ibrida a due stadi	128
8.4.1	Label shifting e definizione del target	128
8.4.2	Stadio 1: Change Point Analysis con baseline personalizzate .	128
8.4.3	Stadio 2: dal Random Forest al RUSBoost	129
8.4.4	Masking della recovery e filtro causale real-time	130
8.5	Protocollo di validazione: Leave-One-Out Cross-Validation	130
8.6	Analisi dei risultati	131
8.6.1	Segnali fisiologici e predizione real-time	131
8.6.2	Matrici di confusione e curva ROC	132
8.6.3	Importanza relativa delle feature	134
8.6.4	Consistenza del modello e robustezza inter-soggetto	135
8.7	Confronto metodologico: modello ASD vs modello piloti	136
8.8	Limiti e prospettive	136
9	Acquisizione sperimentale su soggetti sani tramite braccialetto Empatica EmbracePlus	139

9.1	Il dispositivo Empatica EmbracePlus	139
9.1.1	Protocollo sperimentale	141
9.1.2	Partecipanti	142
9.1.3	Segnali acquisiti	142
9.1.4	Analisi dei segnali per soggetto	143
9.1.5	Problematiche di acquisizione	147
10	Conclusioni	149
10.1	Sintesi del lavoro	149
10.2	Risultati principali	150
10.2.1	Modello ibrido ASD	150
10.2.2	Generalizzazione al dominio aeronautico	151
10.2.3	Acquisizione sperimentale con Empatica EmbracePlus	151
10.3	Contributi metodologici	152
10.4	Limiti del lavoro	153
10.5	Prospettive future	154
	Bibliografia	157
	Ringraziamenti	163

Elenco degli Acronimi

ACC	Accelerometro triassiale
ADI-R	<i>Autism Diagnostic Interview-Revised</i>
ADOS	<i>Autism Diagnostic Observation Schedule</i>
ANS	<i>Autonomic Nervous System</i> (sistema nervoso autonomo)
ANOVA	<i>Analysis of Variance</i> (analisi della varianza)
ASD	<i>Autism Spectrum Disorder</i> (disturbo dello spettro autistico)
AUC	<i>Area Under the Curve</i>
AVNN	<i>Average of Normal-to-Normal intervals</i>
BAS	<i>Behavioral Activation System</i>
BIS	<i>Behavioral Inhibition System</i>
BVP	<i>Blood Volume Pulse</i>
CBCL	<i>Child Behavior Checklist</i>
CPA	<i>Change Point Analysis</i>
DFA	<i>Detrended Fluctuation Analysis</i>
DISC	<i>Diagnostic Interview Schedule for Children</i>
DSM-5	<i>Diagnostic and Statistical Manual of Mental Disorders</i> , Fifth Edition
ECG	Elettrocardiogramma
EDA	<i>Electrodermal Activity</i> (attività elettrodermica)
EDA-R	<i>Electrodermal Activity Response</i>
FC	Frequenza Cardiaca
FIFO	<i>First In, First Out</i>

HF	<i>High Frequency</i> (alta Frequenza, banda HRV 0.15–0.40 Hz)
HR	<i>Heart Rate</i> (Frequenza Cardiaca)
HRV	<i>Heart Rate Variability</i> (variabilità della frequenza cardiaca)
IBI	<i>Inter-Beat Interval</i> (intervallo inter-battito)
LF	<i>Low Frequency</i> (bassa Frequenza, banda HRV 0.04–0.15 Hz)
LOO-CV	<i>Leave-One-Out Cross-Validation</i>
ML	<i>Machine Learning</i> (apprendimento automatico)
MWL	<i>Mental Workload</i> (carico mentale)
ODD	<i>Oppositional Defiant Disorder</i> (disturbo oppositivo provocatorio)
PNS	<i>Parasympathetic Nervous System</i> (sistema nervoso parasimpatico)
PPG	<i>Photoplethysmograph</i> (fotopletismografo)
RF	<i>Random Forest</i>
RMSSD	<i>Root Mean Square of Successive Differences</i>
ROC	<i>Receiver Operating Characteristic</i>
RSA	<i>Respiratory Sinus Arrhythmia</i> (aritmia sinusale respiratoria)
RSA-R	<i>Respiratory Sinus Arrhythmia Response</i>
SCL	<i>Skin Conductance Level</i> (livello tonico dell'attività elettrodermica)
SCR	<i>Skin Conductance Response</i> (risposta fasica dell'attività elettrodermica)
SDNN	<i>Standard Deviation of NN intervals</i>
SHAP	<i>SHapley Additive exPlanations</i>
SIB	<i>Self-Injurious Behavior</i> (comportamento auto-lesivo)
SNA	<i>Autonomic Nervous System</i> (sistema nervoso autonomo)

- SNS** *Sympathetic Nervous System* (sistema nervoso simpatico)
- SNR** *Signal-to-Noise Ratio* (rapporto segnale-rumore)
- SVM** *Support Vector Machine*
- VLF** *Very Low Frequency* (frequenza molto bassa, banda HRV 0.003–0.04 Hz)

Abstract

I comportamenti aggressivi rappresentano una delle manifestazioni più gravi del Disturbo dello Spettro Autistico (ASD), con tassi di prevalenza fino al 56% della popolazione clinica. La letteratura psicofisiologica ha documentato che tali episodi sono preceduti da un'alterazione misurabile del Sistema Nervoso Autonomo, ovvero ritiro del tono vagale e incremento dell'arousal simpatico, rilevabile tramite variabilità della frequenza cardiaca (HRV) e attività elettrodermica (EDA). Il presente lavoro sviluppa un sistema automatico per il rilevamento precoce di questi precursori, acquisiti tramite biosensore da polso (Empatica E4) in contesti di vita quotidiana.

L'architettura proposta è un modello ibrido che integra un algoritmo non supervisionato basato su baseline dinamica e Change Point Analysis, che si adatta continuamente al profilo individuale senza richiedere etichette manuali, e un classificatore supervisionato Random Forest addestrato su dati sintetici, sviluppato per compensare la rarità degli eventi aggressivi nel dataset disponibile. Il modello raggiunge un'AUC di 0.85 su dati artificiali, coerente con i valori standard della letteratura su dati reali (AUC 0.80–0.84). La finestra di predizione ottenuta di circa 2.5 minuti è in linea con il riferimento clinico di 3 minuti adottato come soglia di intervento preventivo.

Il framework è successivamente esplorato in un dominio applicativo secondario, ovvero il monitoraggio del carico mentale nei piloti di aeromobili, per il quale la letteratura recente ha dimostrato l'efficacia delle stesse feature HRV ed EDA. È stata infine condotta una sessione sperimentale su soggetti sani tramite Empatica EmbracePlus, che ha evidenziato criticità sistematiche nel segnale EDA e nel PRV, fornendo indicazioni concrete per il miglioramento del protocollo di acquisizione. L'applicazione dell'architettura su dataset clinici reali di maggiori dimensioni costituisce la prosecuzione naturale di questo lavoro.

1. Introduzione

1.1 Il problema

La gestione delle crisi di aggressività nei soggetti con Disturbo dello Spettro Autistico (ASD) a basso funzionamento rappresenta una delle sfide cliniche più complesse per il personale sanitario, le famiglie e i caregiver [1]. A causa delle gravi limitazioni comunicative, questi individui faticano a esprimere il proprio disagio o l'insorgere di un sovraccarico sensoriale. Di conseguenza, lo stress accumulato sfocia frequentemente in comportamenti problematici o in episodi acuti di disregolazione, spesso definiti in letteratura clinica come *meltdown*.

Attualmente, l'approccio a queste manifestazioni critiche è prevalentemente reattivo e si basa sull'osservazione visiva. I caregiver intervengono nel momento in cui notano i primi segnali comportamentali di agitazione, applicando tecniche di contenimento. Il problema di fondo è che l'intervento avviene quando il soggetto ha già superato la propria soglia di tolleranza: nel momento in cui la crisi risulta visibile all'esterno, il sistema nervoso del paziente risulta già profondamente alterato, rendendo le metodiche di contenimento spesso inefficaci. Per migliorare concretamente la qualità della vita di questi pazienti, è quindi indispensabile passare da una gestione puramente reattiva a una previsione tempestiva dell'evento critico.

Il medesimo principio, ovvero rilevare un deterioramento dello stato autonomico prima che esso si manifesti clinicamente o operativamente, si applica in modo diretto a un secondo dominio di interesse: i piloti di aeromobili. Le statistiche di sicurezza aerea indicano che errori umani correlati a livelli critici di *mental workload* (MWL) contribuiscono a circa il 70% degli incidenti aerei [2]. Anche in questo caso, l'intervento reattivo è insufficiente: un pilota in sovraccarico cognitivo non può segnalare il proprio stato senza interrompere le operazioni di volo e la perdita di consapevolezza avviene in modo silenzioso e progressivo, rendendo impossibile un intervento esterno basato sull'osservazione comportamentale [2]. La necessità di un sistema di monitoraggio continuo e predittivo è dunque comune ai due ambiti, sebbene le popolazioni, i meccanismi fisiologici e le dinamiche temporali siano profondamente diversi.

1.2 I precursori

Le crisi aggressive nei soggetti ASD non compaiono dal nulla [1]. Studi di psicofisiologia mostrano che sono sistematicamente precedute da alterazioni misurabili del Sistema Nervoso Autonomo (ANS) [3][4][5]. Prima che l'aggressività emerga a livello clinico e motorio, l'organismo attiva una risposta fisiologica di preparazione alla reazione di attacco o fuga (*fight-or-flight*), producendo una serie di cambiamenti interni che fungono da veri e propri precursori oggettivi e quantificabili della crisi.

Queste alterazioni coinvolgono due componenti del sistema autonomo: da un lato si registra un incremento dell'attività del sistema nervoso simpatico, frequentemente quantificata tramite l'Attività Elettrodermica (EDA) [5][6]; dall'altro si assiste a una disfunzione del sistema parasimpatico [7][8]. In relazione a quest'ultima componente, la Variabilità della Frequenza Cardiaca (HRV) assume un ruolo determinante. L'HRV è considerato l'indicatore principale del cosiddetto freno vagale, introdotto dalla teoria polivagale di Porges [9]: quando questo freno viene ritirato o risulta cronicamente depresso, la frequenza cardiaca subisce un'accelerazione e il soggetto perde la capacità di auto-regolazione, diventando estremamente vulnerabile agli stimoli ambientali.

Analogamente, anche nello stress operativo dei piloti, i precursori fisiologici precedono la compromissione delle prestazioni. Il concetto di MWL si manifesta attraverso gli stessi mediatori autonomici (aumento dell'HR, riduzione dell'HRV, attivazione fasica dell'EDA) sebbene con dinamiche temporali e ampiezze caratteristiche diverse rispetto alla crisi ASD [10]. In particolare, Rosenblum e colleghi (2025) hanno dimostrato che la risposta cardiaca allo stress acuto non è un fenomeno uniforme, ma si articola in traiettorie temporali distinte: l'HR aumenta immediatamente all'onset del fattore di stress, mentre la riduzione dell'HRV è ritardata e si consolida solo nella fase centrale dell'esposizione [10]. Coelli e colleghi (2026) hanno inoltre identificato l'ampiezza della componente fasica dell'EDA e l'entropia campionaria dell'HRV come le feature più discriminanti tra carico cognitivo elevato e riposo in soggetti sani adulti [11]. Nei piloti, a differenza dei soggetti ASD, la risposta simpatica è tipicamente più contenuta (incremento EDA di +2–4 μ S rispetto a +8–10 μ S nel meltdown) e si sviluppa come un plateau sostenuto di 10–15 minuti, piuttosto che come un picco acuto

di 3 minuti, riflettendo la natura incrementale del carico cognitivo operativo [2][12].

Fino a non molti anni fa, la misurazione di questi parametri richiedeva strumentazioni cliniche invasive. Oggi, la diffusione di dispositivi indossabili dotati di sensori fotoplethysmografici ed elettrodermici consente il monitoraggio continuo dell'ANS in contesti di vita quotidiana [6][13][14]. L'elaborazione in tempo reale di questi parametri permette teoricamente di ricavare una finestra temporale utile per intervenire in modo preventivo, prima che si manifesti l'apice della crisi comportamentale o il picco di degradazione delle prestazioni operative [3][14].

1.3 L'obiettivo e il contesto di ricerca

Il presente lavoro di tesi si pone un duplice obiettivo. Il primo è sviluppare e validare un algoritmo in grado di elaborare i segnali fisiologici acquisiti tramite dispositivi indossabili, al fine di prevedere le crisi di aggressività in soggetti ASD a basso funzionamento. Il secondo obiettivo, che costituisce la componente di generalizzazione del lavoro, è dimostrare la trasferibilità dell'architettura sviluppata, ovvero il modello ibrido CPA+RF, a un dominio diverso: la predizione dello stress operativo nei piloti di aeromobili.

Lo studio si inserisce in un progetto di ricerca più ampio, sviluppato in parallelo con una seconda tesi. Entrambi i lavori condividono la stessa base metodologica per le fasi di acquisizione, filtraggio e sincronizzazione dei dati. I due lavori differiscono però nell'approccio alla modellazione: mentre la tesi parallela ha esplorato l'utilizzo dell'EDA per addestrare un modello supervisionato (Random Forest), il presente elaborato si concentra sull'analisi delle dinamiche cardiovascolari (heart rate e HRV) e sullo sviluppo di una pipeline predittiva non supervisionata.

Questa scelta nasce da una necessità ingegneristica e clinica ben precisa: la fisiologia della popolazione autistica è caratterizzata da fluttuazioni autonome marcate, da un'elevata variabilità sia tra soggetti diversi che all'interno dello stesso soggetto, e da una forte non stazionarietà dei segnali biologici [3][14]. I modelli basati su soglie fisse o su etichette manuali tendono quindi a produrre un numero elevato di falsi positivi. Per ovviare a questo limite, in questo lavoro è stato sviluppato un algoritmo per

il calcolo di una baseline dinamica, capace di aggiornare in tempo reale la stima dello stato fisiologico di riferimento del paziente. A partire da questa baseline è stata applicata la Change Point Analysis (CPA) [15], una tecnica statistica non supervisionata selezionata per la sua capacità di rilevare le discontinuità strutturali del segnale (come il rilascio del freno vagale) senza richiedere un addestramento preliminare basato su annotazioni manuali.

Come fase conclusiva della prima parte del progetto, i due approcci predittivi sono stati integrati in un unico sistema. Utilizzando tecniche di data augmentation per gestire lo sbilanciamento tra le classi, tipico dei dataset clinici, è stato sviluppato e validato un modello ibrido che combina la precisione del modello supervisionato (Random Forest) con la maggiore sensibilità alle variazioni acute propria del modello non supervisionato (CPA) [13][15].

La seconda parte del lavoro estende questa architettura al caso dei piloti. In assenza di un dataset reale di volo etichettato, la generalizzazione è realizzata tramite data augmentation sintetica: i parametri fisiologici dei segnali generati sono calibrati sulla base di regole ricavate dalla letteratura, che quantificano le differenze tra la risposta autonoma nella crisi ASD e nello stress operativo del pilota. In particolare, la baseline dell'EDA nei piloti è attesa nell'intervallo $0.5\text{--}2 \mu\text{S}$ (rispetto a $1\text{--}4 \mu\text{S}$ nei soggetti ASD), il calo dell'RMSSD durante lo stress è graduale ($-20\text{--}30\%$) anziché brusco (-50%), e la durata della fase pre-evento si estende a 10–15 minuti con forma di plateau sostenuto, in contrasto con la rampa brusca di 3 minuti caratteristica del meltdown [2][10][11][12]. Il risultato finale è un sistema di allerta precoce basato su due domini, calibrato sul singolo soggetto e progettato per essere estensibile a qualunque contesto in cui la disregolazione autonoma rappresenti un precursore misurabile di un evento critico.

1.4 Struttura dell'elaborato

Il presente elaborato è strutturato in dieci capitoli. Dopo questa introduzione (capitolo 1), il capitolo 2 presenta una revisione dello stato dell'arte sui precursori fisiologici dell'aggressività nell'ASD, sulle metodologie predittive applicate alla popolazione au-

tistica e, nelle sezioni 2.6 e 2.7, sulle basi fisiologiche dello stress operativo nei piloti e sullo stato dell'arte del rilevamento del MWL in aviazione. Il capitolo 3 descrive il dataset e l'acquisizione dei segnali, illustrando la strumentazione indossabile, il protocollo sperimentale e le caratteristiche delle sessioni di monitoraggio. Il capitolo 4 approfondisce le tecniche di pre-processing applicate ai segnali acquisiti, con particolare attenzione alla decomposizione dell'EDA nelle sue componenti tonica e fasica e alla preparazione del segnale HRV. Il capitolo 5 descrive l'estrazione delle feature e la costruzione del dataset finale.

Il capitolo 6 costituisce il nucleo metodologico della tesi sulla popolazione ASD: vi sono descritti l'implementazione della baseline dinamica individuale, l'applicazione della Change Point Analysis come approccio non supervisionato per il rilevamento dei precursori fisiologici e il modello supervisionato Random Forest soggetto-specifico. Il capitolo 7 presenta il modello ibrido per la predizione delle crisi: ne descrive l'architettura a due stadi (CPA seguita da classificatore Random Forest), la generazione dei dati sintetici, le analisi di robustezza temporale al variare dell'onset della crisi e l'analisi di robustezza al rumore con il relativo intervento correttivo basato sul pre-processing della slope dell'EDA e sulla variabilizzazione dell'ampiezza della rampa sintetica pre-crisi. Il capitolo 8 illustra la generalizzazione dell'architettura ibrida al dominio aeronautico: vi sono descritte la ricalibrazione dei parametri fisiologici per i piloti, la transizione al classificatore RUSBoost per la gestione dello sbilanciamento delle classi, l'introduzione del label shifting a 15 minuti, il masking della recovery e il filtro causale real-time, con validazione tramite Leave-One-Out Cross-Validation su 10 soggetti sintetici.

Il capitolo 9 presenta un'acquisizione sperimentale condotta su due soggetti sani tramite il braccialetto Empatica EmbracePlus, con l'obiettivo di valutare la pipeline di acquisizione e pre-elaborazione in un contesto controllato. I partecipanti hanno indossato il dispositivo durante la visione di film a contenuto emotivo differente, e i segnali esportati dalla piattaforma CareLab, ovvero frequenza cardiaca, PRV-RMSSD, EDA, temperatura cutanea e accelerometria, sono stati analizzati e confrontati con i valori fisiologici di riferimento. Il capitolo discute le criticità emerse, in particolare le anomalie nel segnale EDA e la scarsità di campioni PRV validi, e fornisce indicazioni

operative per il miglioramento del protocollo in sessioni future. Il capitolo 10 chiude il lavoro con le conclusioni, riassumendo i risultati ottenuti nei due domini applicativi e delineando le prospettive future per la validazione clinica e operativa del sistema su dati reali.

2. Stato dell'arte

2.1 Il SNA come base fisiologica della regolazione emotiva nell'ASD

Prima che un comportamento aggressivo emerga a livello motorio, l'organismo attraversa una fase di preparazione gestita dal Sistema Nervoso Autonomo (SNA). Il SNA regola in modo involontario l'attività cardiaca, respiratoria e delle ghiandole sudoripare attraverso due branche con funzioni opposte: il simpatico mobilita le risorse energetiche e accelera il battito cardiaco nella risposta di attacco o fuga (*fight-or-flight*), mentre il parasimpatico, agendo attraverso il nervo vago, favorisce il recupero e abbassa la frequenza cardiaca. Non si tratta di un interruttore on/off: i due sistemi coesistono in un equilibrio continuo, e la qualità di questo equilibrio è ciò che determina la capacità di un individuo di regolare le proprie emozioni.

Il modo più diretto per misurare questo equilibrio è la Variabilità della Frequenza Cardiaca (HRV). Contrariamente a quanto si potrebbe pensare, un cuore sano non batte a intervalli costanti: gli intervalli tra un battito e il successivo (intervalli RR) variano in continuazione, e questa variabilità riflette quanto i due sistemi autonomi stiano interagendo attivamente. La componente ad alta frequenza dell'HRV, nota come aritmia sinusale respiratoria (RSA, *Respiratory Sinus Arrhythmia*), è modulata principalmente dal nervo vago ed è considerata un indice affidabile del tono parasimpatico. Un'HRV alta suggerisce un sistema autonomo reattivo e adattabile; una bassa, invece, indica che il freno vagale è già ritirato e il soggetto si trova in uno stato di attivazione simpatica prolungata, con una capacità di autoregolazione ridotta.

2.1.1 I domini di analisi dell'HRV

La misurazione dell'HRV non produce un unico numero, ma una serie di indici che caratterizzano il segnale su scale temporali e meccanismi fisiologici diversi. Le linee guida della Task Force della Società Europea di Cardiologia [16], integrate dalle indicazioni metodologiche di Berntson e colleghi (1997) [17], distinguono tre domini principali di analisi, ciascuno con implicazioni interpretative specifiche.

Dominio temporale. Gli indici temporali sono calcolati direttamente sulla sequenza degli intervalli RR (o NN, ovvero *Normal-to-Normal*, dopo la rimozione degli ectopici). I più utilizzati sono:

- **SDNN** (*Standard Deviation of NN intervals*, ms): deviazione standard dell'intera serie RR, che riflette la variabilità complessiva includendo le componenti a bassa frequenza legate al tono simpatico. Il suo valore dipende fortemente dalla durata della finestra di acquisizione: su 24 ore cattura la riserva autonoma globale, su finestre di 5 minuti tende invece a riflettere prevalentemente la modulazione parasimpatica.
- **RMSSD** (*Root Mean Square of Successive Differences*, ms): radice quadrata della media dei quadrati delle differenze tra intervalli RR consecutivi. Tra gli indici temporali, è quello più direttamente legato all'attività parasimpatica e alla RSA; mantiene una buona affidabilità anche su finestre brevi (1–2 minuti), il che lo rende preferibile in contesti di monitoraggio real-time [17].
- **pNN50** (%): percentuale di coppie di intervalli RR consecutivi con differenza superiore a 50 ms. Altamente correlato con RMSSD, costituisce anch'esso un marcatore dell'attività vagale, sebbene meno robusto al rumore.
- **Mean HR** (bpm): la frequenza cardiaca media rimane la misura più immediata, ma è fortemente influenzata da fattori come l'attività fisica, la temperatura ambientale e i farmaci. Nei soggetti con ASD va letta rispetto alla baseline individuale del paziente, non confrontata con valori normativi di popolazione [8].

Dominio della frequenza. Applicando la trasformata di Fourier alla serie RR si ottiene una decomposizione spettrale che separa il contributo delle diverse bande di frequenza, ciascuna riconducibile a meccanismi fisiologici distinti [16]:

- **HF** (*High Frequency*, 0.15–0.40 Hz, ms²): cattura la modulazione respiratoria del ritmo cardiaco (RSA) ed è considerato l'indice spettrale più selettivo per l'attività parasimpatica vagale. Per stimarlo in modo affidabile occorre che la

serie RR, dopo ricampionamento uniforme, abbia una frequenza di almeno 1 Hz e che la finestra di analisi copra almeno 1 minuto [16].

- **LF** (*Low Frequency*, 0.04–0.15 Hz, ms^2): banda che rispecchia un contributo misto di attività simpatica e parasimpatica. Il rapporto LF/HF, spesso proposto come indice del bilanciamento simpatico-vagale, produce risultati poco consistenti; nel dominio dell'aviazione, in particolare [2], entrambe le componenti tendono a calare insieme durante lo stress cognitivo elevato, rendendo il rapporto di fatto non interpretabile.
- **VLF** (*Very Low Frequency*, 0.003–0.04 Hz, ms^2): legato a termoregolazione e attività ormonale, richiede finestre di almeno 5 minuti per una stima stabile e risulta di fatto inaccessibile nelle finestre brevi (30–60 s) del monitoraggio in tempo reale [17].

Per il presente lavoro, questo dominio presenta un limite strumentale non trascurabile. Il braccialetto Empatica E4 acquisisce il segnale PPG a 64 Hz, da cui gli IBI vengono estratti e ricampionati uniformemente; le linee guida raccomandano invece almeno 250 Hz sull'ECG grezzo per garantire una localizzazione accurata dei picchi R [16]. A 64 Hz sul segnale BVP, la precisione temporale è ridotta e le fluttuazioni temporali risultanti si propagano come rumore sulle feature spettrali, degradandone il contenuto informativo. Questa considerazione ha orientato la scelta verso gli indici temporali (RMSSD in primo luogo) che mantengono affidabilità anche in presenza di segnale rumoroso.

Indici non lineari. Le serie RR fisiologiche non sono stazionarie e mostrano dinamiche complesse, con correlazioni che si estendono nel tempo e che i metodi lineari non riescono a descrivere adeguatamente [17, 11]. Per questo motivo, la letteratura ha progressivamente introdotto un terzo gruppo di indici:

- **Sample entropy (SampEn)**: quantifica la complessità e l'imprevedibilità della serie RR. Nel lavoro di Coelli et al. [11], condotto su soggetti adulti sani durante task cognitivi, SampEn è risultata una delle feature più discriminanti tra stato

di riposo e attivazione cognitiva nell'analisi SHAP del modello Random Forest, con un aumento significativo durante i task, indipendentemente dal livello di stress indotto.

- **Indici del diagramma di Poincaré (SD1, SD2):** nel piano di Poincaré ogni battito viene rappresentato in funzione del precedente, e la dispersione della nuvola di punti lungo i due assi principali fornisce due misure distinte. SD1, geometricamente equivalente a RMSSD diviso $\sqrt{2}$, cattura la variabilità battito-per-battito, quindi a breve termine; SD2 riflette invece le fluttuazioni a lungo termine. Il loro rapporto SD1/SD2 descrive l'equilibrio tra i due meccanismi di controllo autonomo [17].
- **DFA (Detrended Fluctuation Analysis):** analizza la struttura delle fluttuazioni della serie RR su scale temporali diverse. Il coefficiente α_1 , calcolato su 4–12 battiti, mostra una correlazione con il tono vagale [11] ed è stato proposto come complemento agli indici temporali classici nei protocolli di monitoraggio dello stress.

Nel costruire il classificatore, questi indici sono stati messi da parte per una ragione pratica: per stimarli in modo affidabile servono serie RR piuttosto lunghe, dell'ordine di 200-300 battiti, cioè circa 2-3 minuti a frequenza cardiaca a riposo. Una finestra di questo tipo è compatibile con il dataset ASD disponibile, ma risulta difficile da sostenere in un sistema progettato per funzionare in tempo reale. Per questo motivo, rappresentano più che altro una possibile direzione di sviluppo per versioni future del modello.

2.1.2 La teoria polivagale e il modello di Beauchaine

La teoria polivagale di Porges [9] è il quadro teorico che meglio permette di leggere il ruolo dell'HRV nella regolazione emotiva. Porges descrive il sistema nervoso autonomo come una gerarchia di tre livelli di risposta, ordinati per età evolutiva. Il più recente e sofisticato è il sistema di ingaggio sociale, governato dal ramo ventrale del nervo vago: quando l'ambiente è percepito come sicuro, questo sistema mantiene attivo il freno vagale, tiene bassa la frequenza cardiaca e consente all'individuo di stare in

relazione con gli altri. Quando la minaccia supera questa capacità di contenimento, il freno viene bruscamente tolto: si attiva la risposta simpatica di attacco o fuga, il cuore accelera e l'organismo si prepara all'azione. Se nemmeno questo basta, entra in gioco il terzo livello, il più antico: la risposta di immobilità mediata dal ramo dorsale del nervo vago, un collasso fisiologico che compare nelle situazioni di pericolo estremo e senza via d'uscita.

Un concetto centrale della teoria è la neurocezione: il sistema nervoso valuta continuamente il livello di sicurezza o pericolo dell'ambiente attraverso circuiti subcorticali, al di sotto della soglia della consapevolezza cosciente [9]. Questa valutazione implicita determina quale dei tre livelli venga attivato, indipendentemente dal giudizio razionale del soggetto. Nei bambini con ASD, la neurocezione risulta alterata: stimoli neutri possono essere processati come minacciosi, mentre segnali sociali di sicurezza, come il tono della voce o il contatto visivo, vengono spesso ignorati o male interpretati [9].

Il correlato fisiologico del freno vagale ventrale è misurabile attraverso l'HRV: il tono del ramo vagale mielinizzato si riflette nelle oscillazioni ritmiche della frequenza cardiaca sincronizzate con il respiro, note come aritmia sinusale respiratoria (RSA). Un'HRV elevata indica che il freno vagale è attivo e che il sistema è in uno stato di regolazione flessibile; una riduzione dell'HRV segnala invece il ritiro del freno e l'ingresso in uno stato di attivazione simpatica [18]. Il tono vagale basale e la reattività vagale, ovvero, rispettivamente, il livello di RSA a riposo e la sua variazione in risposta a una domanda ambientale, sono indici distinti del funzionamento autonomico e non vanno trattati come equivalenti [18].

Porges non si concentra tanto su quale livello venga attivato, quanto sulla facilità con cui il sistema riesce a transitare da uno all'altro. Nei soggetti con ASD questa mobilità è ridotta: il sistema nervoso rimane bloccato in stati di iper-attivazione prolungata o, al contrario, di iporesponsività, e stenta a recuperare il tono parasimpatico anche quando la fonte di stress è scomparsa [5]. È proprio questa difficoltà di recupero l'obiettivo del presente lavoro di tesi: la progressiva riduzione dell'HRV nei minuti che precedono un episodio critico costituisce il segnale predittivo su cui si fonda l'intero approccio adottato.

Beauchaine [18] ha ampliato questo quadro teorico integrando la teoria polivagale

con il modello motivazionale di Gray. Ne emerge una visione del sistema nervoso autonomo in cui i rami simpatico e parasimpatico operano in parallelo, ciascuno con un ruolo distinto. Al ramo simpatico appartengono due sistemi motivazionali: il BAS (*Behavioral Activation System*), orientato all'azione e alla ricerca di stimoli, e il BIS (*Behavioral Inhibition System*), che inibisce il comportamento in risposta a segnali di pericolo o novità. Il ramo parasimpatico, attraverso il tono vagale e la sua reattività, governa la capacità di modulare gli stati emotivi in risposta alle richieste dell'ambiente.

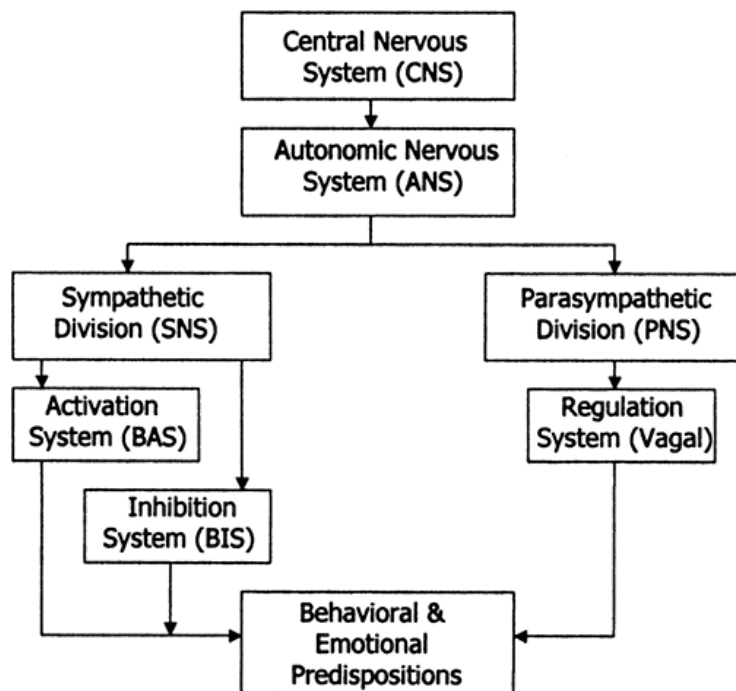


Figura 2.1: Modello integrato del funzionamento del sistema nervoso autonomo proposto da Beauchaine [18]. Le predisposizioni motivazionali (BAS e BIS) ricadono sotto il controllo del sistema nervoso simpatico (SNS); la regolazione emotiva, espressa dal tono vagale e dalla reattività vagale, ricade sotto il controllo parasimpatico (PNS). Entrambi i rami contribuiscono in modo indipendente alle predisposizioni comportamentali ed emotive.

Secondo Beauchaine [18], non è la diagnosi in sé a determinare le difficoltà comportamentali, ma la combinazione specifica di reattività nei tre sistemi. Una RSA ridotta emerge come marcatore *aspecifico* di disregolazione emotiva, trasversale all'aggressività, all'ansia e alla depressione; ciò che distingue questi quadri è il profilo simpatico che vi si accompagna: BAS iperattivo nell'aggressività impulsiva, BIS iperattivo nell'ansia e nella depressione. A questo si aggiunge il ruolo della reattività vagale: un ritiro brusco del freno parasimpatico in risposta a sfide ambientali, indipen-

dentemente dal livello basale di RSA, si associa a labilità emotiva e a episodi esplosivi di panico o aggressione.

La disregolazione autonoma nell'ASD non si presenta in modo uniforme, e questa consapevolezza ha radici lontane. Già negli anni Sessanta, Hutt e colleghi [19] interpretarono stereotipie, iperreattività sensoriale ed evitamento sociale come strategie adattative a uno stato di iper-arousal cronico. Qualche anno dopo, DesLauriers e Carlson [20] avanzarono l'ipotesi opposta: non eccesso, ma deficit di arousal per spiegare i bambini con letargia e scarsa risposta agli stimoli ambientali. Il dibattito con l'integrazione delle due ipotesi: Leekam [21] e Schoen [22] hanno mostrato empiricamente che entrambi i sottogruppi coesistono all'interno della stessa popolazione autistica, uno iper-reattivo, con soglia sensoriale bassa e risposte rapide, l'altro ipo-reattivo, con soglia alta e tempi di risposta allungati. Questa eterogeneità ha una conseguenza metodologica diretta che attraversa tutto il presente lavoro: un classificatore fondato su soglie fisse è destinato a fallire, perché il confine tra risposta normale e anomala non è universale, ma varia da soggetto a soggetto e, nello stesso soggetto, nel corso della giornata.

2.2 Evidenze empiriche sulla risposta cardiovascolare ai fattori di stress nell'ASD

Le prime dimostrazioni empiriche del monitoraggio dell'arousal cardiovascolare nell'ASD vengono dal Groden Center di Providence. Groden e colleghi [7] hanno registrato l'HR in 10 soggetti con ASD esposti a quattro situazioni potenzialmente stressanti: perdita di un gioco, cambio di operatore, tempo non strutturato e consumo di un cibo preferito. Il contributo principale di questo lavoro non stava nei risultati in sé, ma nella dimostrazione che la misurazione era semplicemente fattibile: soggetti con gravi deficit comunicativi tolleravano il dispositivo e producevano dati leggibili. Per quanto riguarda i risultati, il dato più informativo fu che solo 2 partecipanti su 10 mostravano variazioni significative dell'heart rate durante il cambio di operatore, un evento che nella letteratura clinica viene quasi sempre assunto come stressante. Non lo era per otto soggetti su dieci. Un'indicazione chiara che i fattori di stress standardizzati hanno

un valore limitato in questa popolazione e che qualsiasi approccio serio deve orientarsi verso un modello personalizzato.

Vale la pena richiamare, prima di procedere, uno studio precedente che anticipa molti dei temi di Goodwin: Kootz, Marinelli e Cohen [23], nel 1981, confrontarono le risposte di orientamento cardiaco (variazioni della frequenza cardiaca in risposta a stimoli visivi) in bambini con autismo ad alto funzionamento, a basso funzionamento e a sviluppo tipico. I bambini a basso funzionamento partivano da un livello di attivazione basale sistematicamente più alto rispetto agli altri due gruppi, e mostravano di conseguenza una responsività agli stimoli ridotta. La logica è semplice: chi parte già con una frequenza cardiaca elevata ha meno spazio per mostrare un ulteriore aumento chiaramente riconoscibile. Ne deriva un profilo caratterizzato da baseline alta e risposta attenuata, lo stesso schema che Goodwin avrebbe descritto vent'anni dopo. Questo ha un'implicazione importante per i sistemi di allerta precoce: ciò che segnala un possibile rischio non è tanto un valore assoluto elevato, quanto una variazione rispetto alla baseline individuale del soggetto.

Goodwin e colleghi [8] hanno confrontato cinque soggetti con ASD e cinque controlli su sei fattori di stress, con registrazione beat-to-beat della FC tramite un monitor LifeShirt equipaggiato con accelerometro per controllare l'effetto del movimento. Il risultato più sorprendente: il gruppo ASD mostrava risposte cardiovascolari significative solo nel 22% delle esposizioni, contro il 60% dei controlli. Due partecipanti con ASD non risposero a nessuno dei sei fattori di stress, nonostante tutti rispondessero al compito fisico sulla bicicletta ergometrica. Il sistema cardiovascolare funzionava; semplicemente non reagiva agli stressor sociali ed emotivi nel modo atteso.

Gli autori non hanno interpretato questo come ipo-arousal, bensì come chiusura difensiva. I soggetti con ASD presentavano un HR basale di circa 96 bpm contro i 74 bpm dei controlli: uno scarto di 20 bpm che riduce drasticamente lo spazio per rispondere in modo incrementale. A questo si aggiungevano gli effetti carry-over: dopo ciascun fattore di stress l'HR non tornava ai livelli precedenti, rimaneva elevato nelle fasi di riposo successive. Un organismo che non riesce a scaricare l'attivazione tra uno stimolo e l'altro è un organismo che accumula; questo meccanismo offre una spiegazione fisiologica credibile per le escalation comportamentali che si osservano

durante le transizioni.

Entrambi gli studi del Groden Center condividono però un limite strutturale: si basano sulla sola frequenza cardiaca media. Senza misurare l'HRV non si può separare il contributo simpatico da quello parasimpatico, né stimare la flessibilità autonoma del singolo. Due soggetti con identica FC di 96 bpm possono avere profili autonomici opposti: uno con RMSSD basso e HF soppressa (ritiro vagale e dominanza simpatica) l'altro con HRV conservata e attivazione di altra natura. La sola FC non consente di distinguerli, e questa distinzione è esattamente ciò che l'analisi HRV multidimensionale, descritta nella sezione precedente, permette di fare.

Lydon, Healy e Dwyer [4] hanno portato questo lavoro fuori dal laboratorio, registrando la FC di tre bambini con ASD e disabilità intellettiva durante le normali attività scolastiche per circa 20 ore per soggetto. I risultati hanno messo in discussione alcune assunzioni consolidate: la maggior parte dei comportamenti auto-lesivi (SIB) era seguita da un aumento della FC, non da una diminuzione come prevede la *tension reduction hypothesis* [4]. Le stereotipie mostravano invece un pattern diverso: quando comparivano in periodi di bassa attivazione tendevano a portare la FC verso valori medi, suggerendo una funzione di regolazione verso l'alto piuttosto che di scarica. Comportamenti superficialmente simili, dunque, con funzioni autonome opposte: un dato che rende ancora più fragile qualsiasi modello che tratti i comportamenti problema come una categoria unitaria.

2.3 Dall'epidemiologia alla fisiologia: i fattori di rischio per l'aggressività nell'ASD

Kanne e Mazurek [1] hanno realizzato il primo studio sistematico su larga scala sulla prevalenza dell'aggressività nell'ASD, analizzando 1.380 bambini e adolescenti della Simons Simplex Collection con diagnosi confermata tramite ADI-R (*Autism Diagnostic Interview-Revised*) e ADOS (*Autism Diagnostic Observation Schedule*). I numeri sono difficili da ignorare: il 56% dei bambini risultava attualmente aggressivo verso i caregiver, il 32% verso persone non familiari; estendendo l'analisi alle manifestazioni passate, le percentuali salgono al 68% e al 49%. Per dare un termine di paragone, nel-

le popolazioni con disabilità intellettiva senza diagnosi di ASD i tassi di aggressività si attestano tra il 7% e l'11%: uno scarto che esclude l'ipotesi che il fenomeno sia semplicemente una conseguenza del deficit cognitivo.

La parte più rilevante del lavoro riguarda però i predittori. Né il livello cognitivo né la gravità dei sintomi nucleari dell'ASD misurata dall'ADOS risultano associati all'aggressività in modo significativo. I predittori reali sono i comportamenti ripetitivi, ovvero SIB (*Self-Injurious Behavior*), rituali, resistenza al cambiamento. Gravità diagnostica e gravità comportamentale sono quindi due dimensioni che si muovono in modo indipendente, e il fatto che aggressività e comportamenti ripetitivi condividano gli stessi predittori suggerisce qualcosa di più profondo di una semplice co-occorrenza statistica. L'interpretazione più coerente con i dati è che entrambi siano manifestazioni distali dello stesso processo di fondo: la disregolazione del SNA.

A questa lettura, Fenning e colleghi [5] hanno fornito una conferma sperimentale diretta, in quello che è probabilmente lo studio metodologicamente più solido dell'intera letteratura considerata. Su 52 bambini con ASD di età compresa tra 6 e 10 anni, Fenning ha misurato in parallelo la risposta elettrodermica (EDA-R, indicatore simpatico, rilevata con elettrodi Ag-AgCl sul palmo non dominante a 500 Hz) e la risposta dell'aritmia sinusale respiratoria (RSA-R, indicatore parasimpatico, derivata dall'ECG tramite analisi spettrale nelle bande di frequenza respiratoria appropriate per l'età). Entrambe le misure sono state calcolate come variazione rispetto alla baseline durante un compito di sfida cognitiva (la tracciatura di una stella in immagine specchiata), scelto perché in grado di indurre arousal autonomo senza richiedere competenze verbali. Le difficoltà esternalizzanti sono state valutate combinando il Child Behavior Checklist (CBCL) con la sottoscala ODD (*Oppositional Defiant Disorder*) del DISC (*Diagnostic Interview Schedule for Children*); i due punteggi mostravano una correlazione di $r = 0.70$, sufficiente a giustificare la costruzione di un indice composito.

Il risultato centrale è che non esiste un unico profilo autonomico a rischio: ne esistono due, opposti, che portano allo stesso esito. Nel primo (alta EDA-R con forte ritiro della RSA-R) si ha un soggetto iper-reattivo in cui l'attivazione simpatica intensa si combina con un freno vagale insufficiente, e il risultato è una perdita di controllo che sfocia nell'aggressività. Nel secondo (bassa EDA-R con RSA-R ridotta o mode-

rata) si ha un soggetto ipo-reattivo il cui sistema autonomo non risponde agli stimoli, rimanendo bloccato in uno stato di cronico sotto-engagement incapace di autoregolazione. I due profili corrispondono allo squilibrio BAS/BIS descritto da Beauchaine [18]: dominanza del BAS con freno vagale carente nel primo caso, BIS cronico con bassa reattività simpatica nel secondo. Da notare che il campione di Fenning include bambini con disabilità intellettiva (35% del totale, QI tra 47 e 121), il che lo rende più rappresentativo della popolazione clinica reale rispetto a molti lavori costruiti su fasce di funzionamento più omogenee.

A completare il quadro interviene lo studio di Panju e colleghi [24], che hanno misurato l'attivazione simpatica tramite EDA in 27 bambini con ASD e 27 controlli a sviluppo tipico durante un test di stress laboratoriale. I bambini con ASD mostravano un profilo di risposta simpatica atipico, ma l'aspetto più interessante riguarda la relazione con l'ansia misurata dal SCARED (*Screen for Child Anxiety Related Disorders*): nei soggetti con ASD la correlazione tra EDA e ansia era invertita rispetto ai controlli. Un livello elevato di EDA risultava associato a una minore, e non maggiore, ansia percepita dal soggetto. Questo suggerisce che il sistema nervoso autonomo e la consapevolezza emotiva non procedono sempre di pari passo, ma possono essere dissociati.

Dal punto di vista pratico, questo ha una conseguenza importante per il presente lavoro: l'auto-report non è uno strumento affidabile in questa popolazione. Diventa quindi fondamentale utilizzare indicatori oggettivi come l'HRV, che riflettono lo stato autonomo senza dipendere dalla consapevolezza del soggetto, soprattutto quando quest'ultimo ha difficoltà a riconoscere o comunicare il proprio disagio.

2.4 Dalla misurazione alla predizione: dispositivi indossabili e algoritmi automatici

Welch [6] è stato tra i primi a inquadrare il monitoraggio autonomo nell'ASD in chiave ingegneristica, partendo da un'osservazione semplice ma spesso trascurata: I segnali fisiologici sono particolarmente utili in questa popolazione proprio perché non sono sotto il controllo volontario. Parametri come EDA, frequenza cardiaca e HRV

permettono di cogliere lo stato interno del soggetto al di là di ciò che riesce o decide di comunicare. Per questo diventano ancora più importanti quando i canali comunicativi, sia verbali che non verbali, sono limitati o compromessi.

Sul piano metodologico, Kushki e colleghi [13] hanno sviluppato un sistema non supervisionato per il rilevamento automatico dell'arousal in bambini con ASD, basato su un filtro di Kalman applicato alla serie degli intervalli inter-battito (IBI/HR). La logica centrale del loro approccio, stimare dinamicamente la baseline individuale e rilevare in tempo reale le deviazioni da essa, senza alcun addestramento preliminare, è la stessa su cui si fonda l'algoritmo di rilevamento delle anomalie sviluppato nel presente lavoro. La scelta di considerare le variazioni rispetto alla baseline individuale, invece dei valori assoluti, non è un dettaglio tecnico ma una necessità. Nei soggetti con ASD, infatti, le differenze tra individui nella conduttanza basale e nel tono vagale a riposo sono molto più marcate rispetto alla popolazione generale [24]. Di conseguenza, un sistema basato su soglie fisse rischierebbe di non funzionare correttamente per una parte significativa dei pazienti.

I due lavori di Goodwin e colleghi [3, 14] rappresentano il riferimento più diretto per valutare le prestazioni del sistema sviluppato in questa tesi. Il primo, del 2019, ha dimostrato per la prima volta che EDA, HRV e accelerometria acquisiti dall'Empatica E4 (lo stesso dispositivo usato nel presente lavoro) consentono di predire episodi di aggressività in soggetti con ASD, con un AUC di 0.84 nel modello person-dependent, contro 0.71 nel modello globale cross-soggetto. La differenza tra i due non è marginale: addestrare e valutare il classificatore sullo stesso soggetto, usando soglie calcolate sulla distribuzione individuale del segnale, porta a un guadagno di 13 punti percentuali di AUC. Tra le feature più informative nell'ora precedente l'evento figuravano la HR media e gli indici di variabilità a breve termine analoghi a RMSSD. Il lavoro del 2023 [14], condotto su 70 pazienti in 4 ospedali psichiatrici, ha replicato questi risultati su una coorte clinicamente più rappresentativa, raggiungendo un'AUC di 0.80 con un anticipo medio di circa 3 minuti sull'evento aggressivo. È questa finestra di 3 minuti che è stata adottata nel presente lavoro per definire l'intervallo di etichettatura delle classi nel classificatore supervisionato.

In un contesto differente ma metodologicamente affine, Dalmeida e Masala [25]

hanno sviluppato un sistema di rilevamento dello stress nei conducenti, confrontando le feature HRV estratte da ECG e da Apple Watch su 27 soggetti. La selezione tramite correlazione di Pearson, Recursive Feature Elimination ed Extra Trees Classifier ha identificato AVNN, SDNN e RMSSD come variabili più discriminanti; il Random Forest ha raggiunto un recall dell'80%, un valore notevole data la degradazione del segnale tipica dei dispositivi indossabili rispetto all'ECG clinico.

Coelli et al. [11] hanno adottato un approccio multimodale su 62 soggetti sani, combinando ECG, EDA e attività respiratoria durante il Montreal Imaging Stress Task (MIST). I modelli Random Forest con analisi SHAP hanno indicato come feature più discriminanti tra condizioni cognitive stressanti e riposo l'ampiezza della componente fasica dell'EDA, la RSA, la *sample entropy* dell'HRV e il periodo respiratorio medio. Un esito degno di nota è che gli indici HRV lineari (LF, HF, LF/HF) si sono rivelati sensibili alla presenza del task cognitivo in entrambi i gruppi, ma sono stati incapaci di separare i due livelli di stress; le differenze significative emergevano invece sul meanRR e sui parametri EDA. SampEn ha mostrato buona capacità nel distinguere riposo da attivazione cognitiva, ma non nel discriminare tra i due livelli di stress [11]. Per il modello sviluppato in questa tesi, questo suggerisce che un'eventuale estensione agli indici di complessità potrebbe rafforzare la distinzione tra stato basale e stato di attivazione, senza necessariamente migliorare la granularità nella classificazione interna degli stati di stress.

2.5 La Change Point Analysis come strumento per rilevare i precursori

Tutta la letteratura descritta finora porta allo stesso problema metodologico: capire con precisione quando un segnale fisiologico, per quel soggetto, smette di essere nel suo regime fisiologico di riferimento e inizia a indicare una fase di escalation. I metodi classici non sono adatti a questo scopo: l'ANOVA a misure ripetute individua tendenze generali, ma non è pensata per cogliere cambiamenti improvvisi in punti sconosciuti della serie temporale. Allo stesso modo, i modelli di regressione lineare assumono che

il segnale sia stazionario, un'ipotesi che nei dati fisiologici di soggetti con ASD viene spesso violata.

Serve quindi un approccio diverso: uno strumento capace di individuare questi cambiamenti senza dover stabilire in anticipo né quanti saranno né quando si verificheranno.

La Change Point Analysis nasce proprio per rispondere a questo tipo di problema. L'idea di base è individuare i punti in cui le proprietà statistiche di una serie temporale, in particolare media e varianza, subiscono una variazione strutturale, ottimizzando una funzione di costo globale che penalizza l'eccesso di segmentazioni. Esistono diverse famiglie di metodi: approcci esatti che cercano la soluzione ottimale sull'intera serie e metodi bayesiani online che aggiornano la probabilità di cambiamento passo dopo passo. La scelta tra questi dipende dal contesto applicativo e dai vincoli computazionali.

In questo lavoro la Change Point Analysis è stata implementata tramite la funzione `findchangepts` di MATLAB, impostata per rilevare discontinuità nella media dei segnali EDA e HR, con un massimo di due change point per sessione. Questo vincolo non è arbitrario: nelle serie temporali brevi, aumentare il numero di cambiamenti rilevati porta spesso a inseguire il rumore del segnale invece che variazioni fisiologicamente significative. In parallelo, è stata utilizzata una baseline dinamica basata su media mobile, che riduce le fluttuazioni ad alta frequenza preservando invece le variazioni lente e strutturali, quelle che secondo la letteratura precedono le crisi. I dettagli implementativi sono riportati nel capitolo 4.

L'applicabilità della Change Point Analysis ai segnali psicofisiologici longitudinali è stata dimostrata da Rosenfield e colleghi [15] in uno studio sugli attacchi di panico. In undici pazienti con disturbo di panico monitorati per 24 ore, il metodo ha identificato alterazioni cardiorespiratorie significative circa 45 minuti prima dell'insorgenza degli attacchi, mentre tali alterazioni non erano presenti nei periodi di controllo senza eventi.

Il parallelismo con le crisi aggressive nei soggetti con ASD è plausibile: in entrambi i casi si tratta di eventi acuti e relativamente rari, preceduti da una fase di accumulo dell'arousal autonomico che il soggetto non è in grado di comunicare. Se la Change Point Analysis è in grado di intercettare questa fase nel disturbo di panico, lo stes-

so principio può essere esteso, con le opportune modifiche, al contesto analizzato in questa tesi.

2.6 Basi fisiologiche dello stress operativo nei piloti di aeromobili

Estendere il modello sviluppato per le crisi ASD al dominio dell'aviazione civile è possibile perché il meccanismo sottostante è lo stesso: la risposta al *mental workload* (MWL) nei piloti coinvolge lo stesso asse simpatico-vagale descritto nella sezione 2.1, con dinamiche temporali, ampiezze e contesto operativo profondamente diversi. È questa analogia di meccanismo, a parità di fisiologia, con differenza di parametri, che consente un trasferimento cross-dominio calibrato sulla letteratura esistente, invece di una riprogettazione da zero. I limiti di questa analogia sono discussi nella sezione 2.8.

Nel dominio dell'aviazione il costrutto centrale non è lo stress emotivo reattivo tipico dell'ASD, ma il *mental workload* (MWL): il livello di risorse cognitive richieste per soddisfare criteri di prestazione dati, modulato dalle richieste del compito, dal supporto esterno e dall'esperienza del pilota [2]. Le implicazioni operative non sono teoriche: errori umani correlati a livelli critici di MWL sono implicati in circa il 70% degli incidenti aerei. Un caso emblematico è il volo Turkish Airlines TK1951, in cui un pilota non rilevò un'anomalia del sistema di controllo automatico della spinta a causa di un carico cognitivo eccessivo [2]. Non si trattò di imperizia, ma di un sistema cognitivo portato oltre la propria capacità.

A differenza delle crisi nei soggetti con ASD, in cui l'attivazione simpatica tende a essere rapida e concentrata in un intervallo temporale breve, il workload mentale in ambito aeronautico si sviluppa in modo progressivo. Rosenblum e colleghi [10] hanno studiato questa dinamica su 78 partecipanti sani sottoposti al Maastricht Acute Stress Test (MAST), analizzando la risposta in segmenti temporali successivi invece di ridurla a un singolo valore medio.

I risultati mostrano che frequenza cardiaca e variabilità della frequenza cardiaca non seguono lo stesso andamento. La frequenza cardiaca aumenta già nei primi minuti dall'inizio dello stressor, mentre la riduzione dell'HRV è più lenta e diventa signifi-

cattiva solo nella fase centrale dell'esposizione, per poi risalire fino a superare i valori basali nella fase di recupero.

L'analisi tramite modelli a classi latenti (LCMM) ha inoltre identificato un sottogruppo, pari a circa il 15% del campione, con un andamento atipico: in questi soggetti l'HRV aumentava durante l'esposizione prolungata allo stress, e proprio questo gruppo mostrava livelli di ansia più elevati rispetto agli altri [10]. Un risultato che ricorda come la risposta del sistema autonomico non possa essere ridotta a una singola misura riassuntiva.

Un contributo diverso ma complementare viene dallo studio di Yu e colleghi [12], che hanno esposto soggetti sani ad accelerazioni traslazionali passive (PTA) in ascensore, che è una condizione fisicamente analoga alle fasi di decollo e atterraggio. I risultati mostrano un aumento significativo delle risposte elettrodermiche fasiche (SCR) e una riduzione sia di RMSSD sia della potenza HF dell'HRV, confermando l'attivazione della dominanza simpatica in risposta a stimoli vestibolari. Le ampiezze medie per singolo evento (circa $0.35 \mu\text{S}$) e il ritardo di risposta di circa 3.4 secondi dall'onset dell'accelerazione sono stati usati nel presente lavoro come vincoli quantitativi per la costruzione del dataset sintetico nel dominio aviazione, descritto nel capitolo 8.

Calibrazione dei parametri fisiologici per la generalizzazione al dominio aviazione

I lavori analizzati in questo capitolo permettono di ricavare i parametri quantitativi necessari per costruire il dataset sintetico dei piloti, descritto nel capitolo 8. La tabella 2.1 raccoglie le differenze sistematiche tra i segnali della crisi ASD e quelli dello stress acuto del pilota, così come emergono dalla letteratura.

I valori in tabella riflettono due fenomeni fisiologicamente distinti: da un lato la disregolazione emotiva acuta di un soggetto neurodivergente in crisi, dall'altro il sovraccarico cognitivo di un professionista addestrato in assetto operativo. Le sezioni seguenti giustificano ciascun parametro sulla base della letteratura.

EDA baseline ($0.5\text{--}2 \mu\text{S}$) e picco crisi ($+2\text{--}4 \mu\text{S}$). In cabina il corpo del pilota è sostanzialmente fermo: l'attività fisica si riduce a micromovimenti in un am-

Tabella 2.1: Parametri fisiologici della crisi ASD e dello stress acuto nei piloti, derivati dalla letteratura per la calibrazione del dataset sintetico.

Parametro	Crisi ASD	Stress piloti
EDA baseline	1–4 μS	0.5–2 μS
EDA picco evento	+8–10 μS	+2–4 μS
HRV RMSSD baseline	~ 50 ms	~ 55 –70 ms
HRV calo durante stress	–50% (crollo)	–20–30% (graduale)
ACC durante evento	crollo motorio (0.4 \times)	minima variazione
Durata fase pre-evento	3 min (rampa esplosiva)	10–15 min (accumulo lento)
Forma del segnale	picco acuto	plateau sostenuto

biente altamente controllato e il livello tonico della conduttanza cutanea (SCL) parte di conseguenza da valori più bassi rispetto a un soggetto in movimento o con tendenza alla disregolazione sensoriale. Quando compare uno stressor, come un imprevisto, un carico mentale improvviso, o un’accelerazione come quelle studiate da Yu et al. [12], la risposta simpatica fasica si attiva entro 2–6 secondi, ma rimane contenuta in ampiezza. Le misure di Yu et al. [12] indicano ampiezze medie di circa 0.35 μS per singolo evento; valori analoghi emergono dagli indici di AmpSum discussi da Coelli et al. [11]. Il pilota addestrato mantiene il focus operativo anche sotto pressione: il segnale EDA mostra attivazioni frequenti ma moderate (+2–4 μS), ben lontane dalle impennate di +8–10 μS del meltdown autistico, dove la reazione motoria intensa e la sudorazione massiva si sommano alla risposta simpatica pura.

HRV RMSSD baseline (~ 55 –70 ms) e calo durante lo stress (–20–30%). I piloti sono adulti sani sottoposti a controlli medici periodici: una baseline RMSSD di 55–70 ms è fisiologicamente coerente con un buon tono vagale a riposo, in linea con quanto riportato nei modelli di Dalmeida e Masala [25] e nella review di Wang et al. [2]. Sotto carico elevato, come un atterraggio strumentale, o un task cognitivo come il MIST di Coelli et al. [11], il sistema parasimpatico si ritira, ma in modo graduale: non un crollo del 50% come nel meltdown, bensì una riduzione progressiva del 20–30%. Rosenblum e colleghi [10] chiariscono la dinamica temporale: l’HR sale subito, mentre l’HRV si riduce con un ritardo, diventando significativa solo nella fase centrale dell’esposizione.

ACC durante l'evento (minima variazione). Durante una crisi ASD il tracciato dell'accelerometro può essere dominato dall'agitazione psicomotoria o, al contrario, da un collasso motorio brusco. Nel cockpit la situazione è opposta: il pilota può subire carichi cognitivi molto elevati e talvolta forze traslazionali significative [12], ma il corpo rimane ancorato al sedile. L'accelerometro di un wearable registrerà le vibrazioni del velivolo e i movimenti funzionali ai comandi, con fluttuazioni minime rispetto al livello pre-evento.

Durata pre-evento (10–15 min) e forma del segnale (plateau sostenuto). Lo stress del pilota si accumula. Wang et al. [2] e Coelli et al. [11] lo descrivono come un *allostatic load*: richieste attentive, elaborazione continua di dati, decisioni in sequenza, che si stratificano nel tempo senza scaricarsi. Un segmento di avvicinamento e atterraggio può sostenere questo carico per 10–15 minuti, contro la rampa esplosiva di circa 3 minuti che precede una crisi ASD. Rosenblum et al. [10] mostrano che l'attivazione cardiaca, una volta innescata, rimane alta e stabile per tutta la durata del task, con un overshoot di recupero che compare solo a operazione conclusa. È questo plateau prolungato la firma temporale che distingue il sovraccarico del pilota dalla crisi autistica, e che il modello ibrido descritto nel capitolo 7 dovrà imparare a riconoscere.

2.7 Rilevamento del mental workload nei piloti tramite segnali fisiologici

La revisione di Wang e colleghi [2], condotta su 29 studi estratti da PubMed, Scopus e Web of Science nel periodo 2000-2023, è il punto di partenza più utile per orientarsi nello stato dell'arte sul rilevamento del MWL tramite HRV. Il quadro che emerge non è rassicurante: la variabilità metodologica tra gli studi è ampia e i risultati sui cambiamenti degli indici HRV in risposta al carico mentale sono spesso in disaccordo. In generale, l'HR si conferma più reattiva dell'HRV ai cambiamenti di MWL. Diversi lavori osservano variazioni significative della frequenza cardiaca in assenza di qualsiasi differenza rilevabile negli indici di variabilità [2]. Tra gli indici spettrali, la banda HF mostra la maggiore sensibilità alle fluttuazioni a breve termine. Il rapporto LF/HF,

invece, produce risultati contraddittori: quando il carico mentale aumenta, l'attività parasimpatica tende a calare, trascinando con sé sia LF sia HF, e la direzione del cambiamento del rapporto diventa di fatto non interpretabile [2]. È la stessa conclusione a cui si era già arrivati nella sezione 2.1.1: nei contesti di stress cognitivo acuto, RMSSD e SDNN offrono un segnale più stabile e affidabile del rapporto LF/HF.

Sul fronte dei sistemi automatici, gli otto studi che hanno sviluppato modelli di machine learning su HRV mostrano che la classificazione binaria basso/alto MWL è fattibile con risultati promettenti: i valori più alti di precisione e recall raggiungono rispettivamente il 90.88% e il 91.86%, ottenuti con SVM personalizzato su dati individuali [2]. La classificazione multi-classe rappresenta un problema più complesso: quando si passa a tre o quattro livelli di carico, l'accuratezza scende sotto il 50%, e la questione rimane ancora aperta. Un risultato ricorrente in tutta la letteratura riguarda però il ruolo dei modelli soggetto-specifici: i classificatori personalizzati ottengono sistematicamente prestazioni migliori rispetto a quelli generalizzati, a conferma del fatto che la variabilità interindividuale della risposta HRV non può essere trattata come rumore, ma come una caratteristica strutturale del segnale [2].

Questa evidenza motiva, nel presente lavoro, la scelta di calibrare la baseline su ciascun individuo e di adottare un approccio soggetto-specifico nel modello Random Forest, già utilizzato nel dominio ASD ed esteso qui al contesto aeronautico. Un contributo metodologicamente vicino al presente lavoro è quello di Coelli et al. [11], che hanno acquisito in continuo ECG, EDA e attività respiratoria su 62 soggetti sani durante il Montreal Imaging Stress Task (MIST). I modelli Random Forest, valutati tramite analisi SHAP, hanno indicato come feature più discriminanti tra condizioni stressanti e riposo l'ampiezza della componente fasica dell'EDA, la RSA, la *sample entropy* dell'HRV e il periodo respiratorio medio. Un dato che vale la pena sottolineare: le differenze significative tra il gruppo stress e il gruppo controllo emergevano sull'intervallo RR medio e sui parametri EDA, mentre gli indici HRV lineari risultavano sensibili alla presenza del task in entrambi i gruppi, indipendentemente dal livello di stress [11]. La combinazione multimodale di EDA e HRV supera i modelli unimodali, una conclusione coerente con la letteratura più ampia sul rilevamento del MWL in contesti simulati [2], e sostiene la scelta di includere entrambe le modalità come feature primarie nel

dataset sintetico costruito nel capitolo 7.

Un ultimo contributo riguarda un problema pratico ben noto a chi lavora con dati aeronautici: la scarsità di dati etichettati. Hasan e Schiavon [26] hanno analizzato le tecniche di data augmentation applicate a serie temporali di biosegnali, con particolare attenzione ai metodi di generazione parametrica.

I risultati mostrano che una generazione sintetica ancorata a parametri fisiologici realistici riduce il rischio di overfitting rispetto a strategie puramente stocastiche. Inoltre, i modelli addestrati su dati sintetici possono raggiungere prestazioni comparabili a quelli addestrati su dati reali, a condizione che i parametri utilizzati per la generazione derivino da misure sperimentali.

È proprio questo l'approccio adottato nel capitolo 8 per la costruzione del dataset dei piloti.

2.8 Posizionamento del presente lavoro nella letteratura

La letteratura descritta in questo capitolo traccia un quadro abbastanza definito nei contorni generali, anche se con aree ancora aperte sul piano applicativo. Le crisi aggressive nell'ASD sono precedute da alterazioni del SNA misurabili oggettivamente [3, 5, 4]; frequenza cardiaca e HRV sono indicatori affidabili di queste alterazioni [8, 7]; i dispositivi indossabili consentono la misurazione continua e non invasiva in contesti reali [6, 13]; e algoritmi automatici applicati a questi segnali possono raggiungere prestazioni predittive clinicamente utili, con un anticipo di 2–3 minuti sull'evento [14].

Restano tuttavia due problemi principali ancora aperti. Il primo riguarda l'eterogeneità fisiologica: la baseline di un soggetto con ASD non è stabile, ma varia nel corso della giornata, tra sessioni diverse e in funzione di fattori contestuali difficili da controllare. In queste condizioni, un sistema basato su soglie fisse non è in grado di gestire la variabilità tipica della pratica clinica reale [7, 13].

Il secondo problema è la rarità degli eventi: le crisi aggressive conclamate sono troppo poco frequenti per consentire l'addestramento di un classificatore supervisionato esclusivamente su dati reali, senza ricorrere a tecniche di bilanciamento o a dati

sintetici. Anche il dataset utilizzato in questo lavoro riflette questa limitazione: nelle sessioni disponibili non si è verificato alcun episodio di crisi aggressiva grave. Questo lavoro affronta in modo esplicito entrambi i problemi. Il primo, legato all'eterogeneità fisiologica, viene risolto tramite un algoritmo non supervisionato basato su baseline dinamica e Change Point Analysis, ispirato rispettivamente a Kushki et al. [13] e Rosenfield et al. [15]. Questo approccio consente di adattare continuamente la stima dello stato basale al singolo soggetto in tempo reale, evitando la necessità di una fase di calibrazione offline.

Il secondo problema, quello della rarità degli eventi, viene invece affrontato attraverso un modello ibrido che integra la Change Point Analysis con un classificatore Random Forest addestrato su dati sintetici generati mediante tecniche di data augmentation. In questo modo è possibile produrre campioni della classe minoritaria coerenti con i parametri fisiologici riportati in letteratura.

La tesi si colloca in questo contesto estendendo il modello ibrido CPA+RF al dominio dell'aviazione civile. Il trasferimento è giustificato dalla comune base fisiologica del sistema simpatico-vagale, ma presenta anche limiti specifici che è opportuno esplicitare. Dal punto di vista del segnale, la crisi nei soggetti con ASD è caratterizzata da un trigger improvviso, un'ampiezza elevata e una componente motoria dominante nell'accelerometro; il workload mentale del pilota, invece, si sviluppa in modo graduale, con ampiezze più contenute e una componente motoria quasi assente.

I parametri del dataset sintetico aeronautico sono stati calibrati su queste differenze, come descritto nella sezione precedente e nel capitolo 8. Non si tratta quindi di applicare lo stesso modello a un dominio diverso, ma di utilizzare la stessa architettura adattandone i parametri alla letteratura di riferimento. In assenza di dati reali etichettati di piloti per una validazione diretta, questa rimane un'estensione metodologicamente solida, ma ancora aperta sul piano empirico.

3. Dataset e acquisizione

3.1 Il campione di studio e il contesto sperimentale

I dati analizzati nel presente elaborato provengono da un protocollo osservazionale dedicato al monitoraggio fisiologico continuo di individui con Disturbo dello Spettro Autistico (ASD). Le acquisizioni si sono svolte nel novembre 2022, in due giornate distinte presso una struttura terapeutica specializzata nell'intervento con questa popolazione. Il campione è composto da tre soggetti, due di sesso maschile e uno di sesso femminile, con diagnosi di ASD.

Data la dimensione ridotta del campione e la complessità del quadro clinico, il presente lavoro si configura come uno studio pilota. L'obiettivo non è la validazione statistica su larga scala, bensì verificare la fattibilità e la robustezza di una pipeline di acquisizione e pre-processing in un contesto clinico reale, in cui le condizioni operative non sono controllabili con la stessa precisione tipica di un ambiente di laboratorio.

I partecipanti sono identificati nel testo tramite i codici soggetto 1, soggetto 2 e soggetto 3. In conformità con la normativa vigente sul trattamento dei dati personali sensibili in ambito sanitario, i dati anagrafici non sono stati inclusi nella presente trattazione. Le informazioni relative agli eventi clinici occorsi durante le sessioni sono riportate in forma anonimizzata, esclusivamente nei casi in cui risultino rilevanti per l'interpretazione dei segnali fisiologici.

Le acquisizioni si sono svolte all'interno di una stanza multisensoriale attrezzata con materiali e stimoli di diversa natura: palline luminose a variazione cromatica, fibre ottiche, clessidre fluorescenti, materiale fluorescente con luce nera di Wood, proiezioni ambientali su soffitto e pavimento e un letto vibroacustico. Era inoltre presente un materasso physioform utilizzato per la stimolazione posturale e somatica. Il protocollo non ha previsto una fase di resting state statico: i soggetti hanno indossato il dispositivo durante le normali attività terapeutiche in corso nella stanza, che comprendevano l'esplorazione libera degli oggetti sensoriali, la stimolazione tattile, di pressione e vibratoria sul corpo e l'interazione con gli operatori presenti. Il braccialetto è stato

trasferito sequenzialmente tra i soggetti nel corso di ciascuna giornata di acquisizione, secondo una tempistica registrata nel protocollo cartaceo.

La scelta di monitorare HR e HRV in questa popolazione non è casuale, ma motivata da ragioni fisiologiche precise. Come discusso nel capitolo 2, i soggetti con ASD presentano frequentemente gravi limitazioni nella comunicazione verbale del proprio stato emotivo e di allerta [13]: non è possibile chiedere loro di riferire come si sentono nel corso della sessione, né interpretare l'assenza di segnali esteriori come assenza di disagio interno. HR e HRV permettono di aggirare questo limite, poiché riflettono l'equilibrio simpatico-vagale descritto dalla teoria polivagale di Porges [9] in modo del tutto indipendente dalla consapevolezza e dalla volontà del soggetto. In particolare, l'indice RMSSD, derivabile dalla sequenza degli IBI acquisita dal sensore PPG, è considerato il marcatore temporale più affidabile del tono vagale ed è calcolabile anche su finestre brevi di 1–2 minuti [17], compatibili con la durata delle sessioni disponibili. Il profilo di HR e HRV di ciascun soggetto verrà utilizzato nel capitolo 5 come insieme di feature primarie per il classificatore supervisionato.

3.2 Strumentazione di acquisizione

Per la raccolta dei segnali fisiologici è stato utilizzato il dispositivo indossabile Empatica E4 (Empatica Inc., Milano, Italia), un braccialetto progettato per l'acquisizione continua e non invasiva di segnali fisiologici multimodali in contesti di vita reale [27]. Il dispositivo ha dimensioni contenute ($44 \times 40 \times 16$ mm) e pesa 25 g, caratteristiche che lo rendono tollerabile anche per soggetti con ipersensibilità tattile, frequente nella popolazione con ASD [21]. I dati vengono memorizzati internamente per un massimo di 36 ore e successivamente trasferiti in formato .csv tramite connessione USB o bluetooth, rendendoli disponibili per le fasi di elaborazione successive. Il medesimo dispositivo è stato impiegato nei lavori di riferimento di Goodwin e colleghi [3, 14], il che rende il confronto diretto tra i risultati metodologicamente idoneo.

L'Empatica E4 integra quattro sensori, descritti di seguito insieme alla giustificazione della loro rilevanza per il presente lavoro.

Sensore fotopletismografico (PPG). Il sensore PPG rileva le variazioni del vo-



Figura 3.1: Architettura del sistema Empatica E4: il dispositivo (1) trasmette i segnali via bluetooth (2) all'applicazione mobile, che li carica sul cloud Empatica Connect (3) per la successiva esportazione e analisi. Immagine tratta da Empatica Inc.

lume ematico nel microcircolo periferico tramite due LED verdi e due LED rossi posizionati sulla faccia inferiore del dispositivo a contatto con il polso. Il segnale prodotto è il Blood Volume Pulse (BVP), campionato a 64 Hz, da cui vengono derivati la frequenza cardiaca (HR) e la sequenza degli intervalli inter-battito (IBI) [3]. La derivazione degli IBI avviene in due fasi: il rilevamento del picco sistolico nell'onda pletismografica e il calcolo della differenza temporale tra picchi consecutivi. La precisione di questo processo dipende direttamente dalla frequenza di campionamento: a 64 Hz, la risoluzione temporale minima tra due campioni è pari a $1/64 \approx 15.6$ ms, il che introduce un errore di quantizzazione sui tempi dei picchi che si propaga sugli IBI e, di conseguenza, su tutte le feature HRV derivate. Le linee guida della Task Force della Società Europea di Cardiologia raccomandano frequenze di campionamento non inferiori a 250 Hz sull'ECG grezzo [16], mentre Berntson e colleghi (1997) specificano che frequenze inferiori degradano in modo significativo la stima degli indici frequenziali (LF, HF) e, in misura minore, di quelli temporali [17]. Per questo motivo, nel presente lavoro si è scelto di privilegiare gli indici temporali a breve termine, in particolare RMSSD e mean HR, rispetto agli indici spettrali LF/HF, la cui stima affidabile richiede sia una frequenza di campionamento adeguata sia finestre di analisi di almeno 5 minuti [16]. L'elaborazione degli IBI ha inoltre richiesto algoritmi di interpolazione dedicati per il ricampionamento uniforme della serie RR [17], descritti nel capitolo successivo.

Sensore di attività elettrodermica (EDA). Il sensore EDA misura le variazioni della conduttanza cutanea tramite due elettrodi in acciaio inossidabile posizionati sulla fascia del braccialetto, a contatto con la faccia interna del polso. Il campionamento avviene a 4 Hz, con una risoluzione di circa 900 pSiemens e un range operativo compreso tra 0.01 e 100 μ Siemens [27]. Il segnale fornisce una misura diretta dell'attivazione del sistema nervoso simpatico attraverso l'innervazione delle ghiandole sudoripare eccrine: ogni incremento dell'arousal simpatico produce un aumento della conduttanza misurabile entro pochi secondi. Rispetto ai siti di misura standard come le falangi o il palmo, il polso presenta una densità inferiore di ghiandole sudoripare eccrine, il che può attenuare l'ampiezza assoluta del segnale rilevato [27]. Nel contesto del presente lavoro questa limitazione è considerata accettabile: l'analisi non si basa su valori assoluti di conduttanza, bensì sulle variazioni relative rispetto alla baseline individuale, che risultano clinicamente informative indipendentemente dall'ampiezza assoluta.

Accelerometro triassiale (ACC). L'accelerometro registra le accelerazioni del polso lungo i tre assi spaziali (X, Y, Z) a 32 Hz, con un range predefinito di ± 2 g [27]. Nel contesto del presente lavoro il segnale ACC assume un ruolo ausiliario: viene utilizzato per identificare le finestre temporali contaminate da artefatti da movimento, nelle quali la stima degli IBI dal segnale PPG risulta inaffidabile e deve essere corretta o scartata prima del calcolo delle feature HRV [3].

Termometro a infrarossi (TEMP). Il sensore misura la temperatura cutanea superficiale tramite un termopila a infrarossi campionata a 4 Hz, con una risoluzione di 0.02 °C e un'accuratezza di ± 0.2 °C nell'intervallo fisiologico 36–39 °C [27]. La temperatura cutanea è un indice della vasocostrizione periferica mediata dal sistema simpatico: in condizioni di arousal elevato, la ridistribuzione del flusso sanguigno verso i muscoli produce un calo della temperatura cutanea alle estremità. Nel presente elaborato il segnale TEMP non è stato incluso tra le feature del classificatore per tre ragioni concorrenti. In primo luogo, la temperatura cutanea presenta una latenza fisiologica significativamente mag-

giore rispetto a EDA e HR: la vasocostrizione simpatica periferica si traduce in una variazione misurabile al polso nell'ordine dei minuti, non dei secondi [16], rendendola uno strumento inadeguato per la rilevazione precoce di eventi acuti nell'orizzonte temporale di 1–3 minuti richiesto da un sistema di allerta. In secondo luogo, il segnale è fortemente influenzato da fattori confondenti non autonomici: la temperatura ambiente della stanza, il movimento del braccialetto sul polso e il calore prodotto dall'attrito meccanico durante lo stimming motorio introducono variazioni che si sovrappongono al segnale di origine simpatica e ne compromettono l'interpretabilità clinica in assenza di un controllo ambientale rigoroso. In terzo luogo, il trasferimento sequenziale del dispositivo tra i tre soggetti nel corso della stessa giornata implica che il sensore non avesse raggiunto l'equilibrio termico con il polso di ciascun nuovo soggetto prima dell'inizio della sessione [27], introducendo un errore nei primi minuti di ogni acquisizione. Il segnale TEMP è stato comunque caricato e conservato nella pipeline per eventuali sviluppi futuri.

3.3 Protocollo e sessioni sperimentali

L'analisi si basa sui dati acquisiti nelle sessioni del 16 e del 18 novembre 2022. In entrambe le giornate il dispositivo è stato indossato in modo sequenziale: al termine della sessione di un soggetto il braccialetto veniva rimosso e trasferito al successivo, seguendo il piano terapeutico stabilito. Gli orari di inizio e fine di ciascuna sessione sono stati registrati manualmente dall'operatore nel protocollo cartaceo. Queste annotazioni costituiscono l'unico riferimento temporale disponibile per delimitare le sessioni all'interno del file di dati continuo prodotto dal sensore e sono state utilizzate per allineare i tracciati fisiologici agli eventi osservati in stanza. Questa procedura introduce inevitabilmente una certa incertezza nell'identificazione precisa degli eventi, stimabile nell'ordine di qualche secondo, di cui si è tenuto conto nella fase di analisi.

3.3.1 Acquisizioni del 16 novembre 2022

Il monitoraggio ha avuto inizio alle 16:20 con una sessione relativa al soggetto 2, che ha interagito liberamente con oggetti luminosi a variazione cromatica. Alle 17:10, dopo 50 minuti, il dispositivo è stato trasferito al soggetto 1. Durante questa sessione, della durata di 36 minuti, il soggetto 1 ha sperimentato pressione, stimolazione uditiva e vibratoria, interazione con fibre ottiche e clessidre fluorescenti. Entrambi i soggetti hanno assistito a proiezioni ambientali sul soffitto e sul pavimento e hanno interagito con gli operatori scambiandosi oggetti luminosi.

Alle 17:46 il dispositivo è stato applicato al soggetto 3, immediatamente dopo un episodio acuto di disregolazione comportamentale trattato con intervento farmacologico in estemporanea. Il soggetto 3 ha indossato il braccialetto fino alle 18:42, per una durata totale di 56 minuti. La registrazione di questa sessione è di particolare interesse per l'analisi dell'HRV: l'intervento farmacologico e la fase di recupero successiva all'episodio di disregolazione dovrebbero riflettersi in una progressiva normalizzazione del profilo HR e in un graduale recupero del tono vagale, rilevabile come incremento di RMSSD nel tempo.

3.3.2 Acquisizioni del 18 novembre 2022

Il soggetto 3 ha indossato il dispositivo a partire dalle 10:20, durante il tragitto verso la stanza. L'inizio effettivo delle attività è avvenuto alle 10:35, con una sessione di lavoro della durata di circa 45 minuti. La sessione ha incluso contatto di pressione sulle spalle e stimolazione somatica primaria e di pressione sul resto del corpo. Il dispositivo è stato rimosso alle 11:20, per un tempo totale di indosso di 60 minuti.

Alle 11:20 il dispositivo è stato trasferito al soggetto 2, che per 25 minuti ha sperimentato materiali fluorescenti, fibre ottiche e luce di Wood.

Alle 11:46 il dispositivo è stato applicato al soggetto 1, sdraiato sul letto vibroacustico. Il soggetto ha interagito con le fibre ottiche e ha sperimentato il movimento del letto in diverse modalità: con l'operatore che lo muoveva in prossimità del petto, dei piedi e infine contemporaneamente in entrambe le zone.

3.3.3 Dal protocollo cartaceo al protocollo digitale

Le annotazioni scritte nel protocollo cartaceo durante le sessioni sono state tradotte in una struttura dati MATLAB denominata `events`, generata dallo script `protocollo.m` e salvata nel file binario `protocollo.mat`. Questo passaggio costituisce il primo stadio della pipeline computazionale e ha richiesto una serie di scelte metodologiche esplicite, descritte nelle sezioni successive.

Definizione dei timestamp di sessione

Per ciascuna delle sei sessioni analizzate, il timestamp di inizio e di fine è stato derivato direttamente dagli orari annotati nel registro cartaceo, con tre accorgimenti:

- **Esclusione del tempo di transito:** nella sessione del soggetto 3 del 18 novembre, il braccialetto è stato applicato alle 10:20 durante il tragitto verso la stanza terapeutica. Il timestamp di inizio sessione è stato tuttavia fissato alle 10:35, corrispondente all'ingresso effettivo in stanza, escludendo i 15 minuti di cammino durante i quali i segnali fisiologici rifletterebbero l'attività motoria del trasporto e non quella terapeutica di interesse.
- **Margini di cambio sensore:** tra la rimozione del dispositivo da un soggetto e la sua applicazione al successivo intercorre un intervallo di 1–5 minuti, visibile nei gap tra i timestamp di fine e inizio degli eventi consecutivi (ad es. 17:10–17:15 tra soggetto 2 e soggetto 1 il 16 novembre, e 11:45–11:47 tra soggetto 2 e soggetto 1 il 18 novembre). Questi margini sono stati introdotti deliberatamente per escludere la fase iniziale di adattamento del segnale EDA e PPG successiva all'applicazione del dispositivo.
- **Sincronizzazione al fuso orario locale:** tutti gli orari delle sessioni sono stati impostati con fuso orario Europe/Rome. Questa scelta è necessaria perché il sensore salva gli orari dei campioni come conteggio di secondi a partire dal 1° gennaio 1970 (formato standard nei sistemi informatici, indipendente dal fuso orario), e la funzione `load_empatica` si occupa di convertire questo numero grezzo in orari leggibili nel fuso locale. Senza questa conversione, gli orari

del sensore e le annotazioni scritte a mano nel protocollo cartaceo risulterebbero sfasati di un'ora, rendendo impossibile l'allineamento corretto tra i tracciati fisiologici e gli eventi osservati in stanza.

Codifica delle etichette di classe

A ciascuna sessione è stata assegnata un'etichetta numerica (label) che riflette il livello di arousal atteso e la valenza dell'attività svolta. La codifica è ternaria:

- **Classe 0 (relax/sensoriale):** sessioni a basso arousal con valenza positiva, in cui il soggetto esplorava liberamente gli oggetti sensoriali o era sottoposto a stimolazione vibroacustica passiva.
- **Classe 1 (cognitive load/work):** sessioni a medio arousal con valenza positiva, corrispondenti a stimolazione somatica strutturata con contatto fisico guidato dall'operatore.
- **Classe 2 (stress/recovery):** sessioni ad alto arousal con valenza negativa, corrispondenti alla fase di recupero successiva a un episodio acuto con intervento farmacologico.

La distribuzione delle etichette risultante è sbilanciata: 4 sessioni appartengono alla classe 0, 1 alla classe 1 e 1 alla classe 2. Uno squilibrio di questo tipo era inevitabile dato il carattere osservazionale del protocollo: gli eventi ad alto arousal non possono essere né pianificati né replicati a piacimento in una popolazione clinica.

Organizzazione dei file grezzi

I file CSV prodotti dall'Empatica E4 sono organizzati in due cartelle distinte, una per ciascuna giornata di acquisizione. Il nome di ogni cartella corrisponde al momento esatto in cui il dispositivo è stato acceso quella mattina, espresso come numero di secondi dal 1° gennaio 1970 (1668588017_A02724 per il 16 novembre e 1668763066_A02724 per il 18 novembre). Poiché il braccialetto non è mai stato spento tra una sessione e l'altra nel corso della stessa giornata, tutti e tre i soggetti condividono lo stesso file CSV: non esiste un file separato per soggetto 1, uno per soggetto 2

e uno per soggetto 3, ma un'unica registrazione continua che copre l'intera giornata. L'unico modo per isolare i dati di ciascun soggetto è ritagliare la porzione di traccia compresa tra l'orario in cui il braccialetto è stato indossato e quello in cui è stato rimosso, secondo le annotazioni del protocollo cartaceo.

3.3.4 Riepilogo delle sessioni analizzate

La tabella 3.1 riassume le sei sessioni che costituiscono il dataset analizzato, con i parametri di acquisizione derivati direttamente dagli intervalli temporali definiti nel protocollo digitale e dalle frequenze di campionamento del dispositivo. I campioni teorici sono calcolati come $N = f_s \times \Delta t$, dove Δt è la durata della sessione in secondi; la colonna HRV riporta il valore di RMSSD calcolato sull'intera sessione dalla pipeline di pre-processing. Il segnale IBI non è campionato a frequenza fissa: il numero di eventi dipende dalla frequenza cardiaca istantanea ed è variabile per definizione.

Tabella 3.1: Riepilogo delle sei sessioni analizzate. I campioni EDA, HR e ACC sono valori teorici calcolati dalla durata della sessione e dalla frequenza di campionamento del dispositivo. Il RMSSD è calcolato sull'intera sessione dalla funzione `calc_hrv`.

Sogg.	Sessione	Classe	Durata [min]	EDA [camp.]	HR [camp.]	ACC [camp.]	RMSSD [ms]
<i>16 novembre 2022</i>							
Sogg. 2	Baseline sensoriale	0	50	12 000	3 000	96 000	53
Sogg. 1	Stimolaz. sensoriale	0	31	7 440	1 860	59 520	97
Sogg. 3	Recovery post-crisi	2	55	13 200	3 300	105 600	91
<i>18 novembre 2022</i>							
Sogg. 3	Work / pressione prof.	1	45	10 800	2 700	86 400	74
Sogg. 2	Stimolaz. sensoriale	0	24	5 760	1 440	46 080	43
Sogg. 1	Vibroacustica	0	13	3 120	780	24 960	92

Come si vede nella tabella, la durata delle sessioni varia in modo considerevole: si va dai 13 minuti della sessione vibroacustica del soggetto 1 ai 55 minuti della sessione di recovery del soggetto 3. Non si tratta di una scelta progettuale, ma di una conseguenza diretta del fatto che i tempi di indosso erano decisi dall'operatore in base

al piano terapeutico della giornata. Le implicazioni pratiche di questa variabilità sulla definizione delle finestre di analisi e sul numero di feature estratte per sessione sono discusse nel capitolo 5.

3.4 Criticità nell'acquisizione dei dati

I dati raccolti presentano le difficoltà tipiche delle misurazioni fisiologiche condotte in contesti reali con soggetti con ASD: l'attività motoria spontanea e i comportamenti imprevedibili costituiscono fonti sistematiche di rumore e artefatti da movimento [4, 3].

La sessione vibroacustica del soggetto 1 ha rappresentato il caso più critico. Le vibrazioni continue del letto hanno introdotto micro-movimenti costanti sul sito di misurazione, causando una frammentazione nella lettura del sensore PPG. Le conseguenze sulla qualità delle feature HRV sono dirette: nelle finestre temporali in cui il segnale PPG è disturbato, la serie degli IBI risulta incompleta o presenta picchi mal localizzati, rendendo i valori di RMSSD e mean HR inaffidabili. Le finestre con una quota rilevante di IBI mancanti o anomali sono state escluse dall'analisi, in accordo con le raccomandazioni metodologiche per l'analisi della variabilità cardiaca [17, 16]. Il risultato pratico è una riduzione della quantità di dati utilizzabili per questo soggetto nelle fasi di estrazione delle feature descritte nel capitolo 5.

Per il soggetto 2 il problema si è manifestato in modo diverso. I valori basali di conduttanza cutanea erano molto ridotti, al limite del rumore strumentale, rendendo il segnale EDA scarsamente informativo per buona parte della sessione. Le cause sono probabilmente molteplici e difficili da separare: possibili problemi di aderenza degli elettrodi a secco, variazioni di impedenza cutanea locale, i limiti già discussi del posizionamento al polso [27] e un profilo individuale di bassa responsività elettrodermica [5]. In questo caso è stato il segnale HRV derivato dal PPG a risultare qualitativamente più stabile, confermando l'utilità di un approccio multimodale: nei soggetti con bassa responsività elettrodermica, HR e HRV conservano un contenuto informativo che l'EDA da sola non sarebbe in grado di fornire.

Queste criticità rendono evidente che i segnali grezzi non possono essere utilizzati

direttamente all'interno di un classificatore. Prima di estrarre qualsiasi feature è necessario sottoporre i dati a una pipeline di filtraggio, correzione e sincronizzazione. Le metodologie adottate per questa fase di pre-processing, con particolare attenzione alla correzione degli IBI anomali e al ricampionamento della serie RR, sono descritte nel capitolo successivo.

4. Pre-processing

Il pre-processing rappresenta una fase critica di qualsiasi pipeline di analisi dei segnali fisiologici, a maggior ragione quando i dati provengono da acquisizioni condotte al di fuori del laboratorio su soggetti con comportamento motorio imprevedibile. I segnali grezzi ottenuti dall'Empatica E4 non sono direttamente utilizzabili per l'estrazione di feature significative: prima di qualsiasi analisi è necessario affrontare il problema degli artefatti da movimento e delle irregolarità nella qualità del segnale. Come discusso nel capitolo precedente, le sessioni acquisite presentano criticità eterogenee tra i soggetti, che rendono necessario un approccio di elaborazione robusto e adattivo. Questo capitolo descrive le scelte metodologiche adottate per ciascuna di queste problematiche, seguendo l'ordine logico della pipeline implementata in MATLAB.

4.1 Caricamento e segmentazione

I file prodotti dall'Empatica E4 in formato `.csv` sono stati caricati tramite una funzione dedicata (`load_empatica`), che estrae il timestamp di inizio acquisizione e la frequenza di campionamento dalle prime due righe di ciascun file e ricostruisce un vettore temporale assoluto sincronizzato al fuso orario Europe/Rome. La funzione restituisce tre valori: il vettore dei campioni grezzi, il vettore temporale e la frequenza di campionamento (f_s); quest'ultima è indispensabile per tutte le operazioni di filtraggio e segmentazione successive. Questa struttura è comune a tutti i segnali campionati a frequenza fissa: EDA (4 Hz), ACC (32 Hz), frequenza cardiaca HR (1 Hz) e temperatura cutanea TEMP (4 Hz).

Il file `IBI.csv` è strutturalmente diverso dagli altri segnali. Mentre i file a frequenza fissa contengono sequenze di campioni equispaziati nel tempo, il file `IBI` registra gli eventi di battito cardiaco rilevati dal sensore PPG: ogni riga contiene il tempo trascorso dall'inizio dell'acquisizione e la durata dell'intervallo fino al battito successivo, espressa in secondi. Il dispositivo stesso esegue il rilevamento dei picchi sistolici e un primo filtraggio degli artefatti prima di registrare questi valori [3]. Ne consegue che il file `IBI` non possiede una frequenza di campionamento fissa, e la funzione dedicata

`load_empatica_ibi` restituisce pertanto soltanto due valori: il vettore degli intervalli grezzi e il vettore dei relativi timestamp.

Questa sequenza temporale di intervalli tra battiti consecutivi, nota in letteratura come tachogramma, è un segnale a campionamento asincrono: i campioni vengono quindi rappresentati in visualizzazione come punti isolati anziché come traccia continua, poiché ciascun punto corrisponde a un evento di battito e non a un istante di campionamento regolare. Questa caratteristica spiega perché l'elaborazione degli IBI richieda algoritmi di interpolazione dedicati prima di poter essere analizzata nel dominio della frequenza per il calcolo dell'HRV [16, 17].

È stato caricato anche il segnale di temperatura cutanea (`TEMP.csv`, 4 Hz), il cui utilizzo è rimandato a sviluppi futuri della pipeline.

Struttura del protocollo e gestione degli eventi

La pipeline opera su una struttura dati denominata `events`, caricata dal file `protocollo.mat` generato dallo script `protocollo.m`. Ciascuna delle sei sessioni di acquisizione è codificata come un elemento della struttura con i seguenti campi: un identificativo univoco (`id`), il nome del soggetto (`soggetto`), un'etichetta di classe `target` (`label`), il percorso alla cartella dei file grezzi (`folder`) e i timestamp di inizio e fine della sessione (`start`, `end`), entrambi di tipo `datetime` con fuso orario Europe/Rome. Le etichette adottano la seguente codifica: 0 per le sessioni a basso arousal (relax/sensoriale), 1 per le sessioni a medio arousal con valenza positiva (cognitive load/work) e 2 per le sessioni ad alto arousal con valenza negativa (stress/recovery).

I sei eventi vengono elaborati in sequenza. Per evitare che l'intero processo si interrompa a causa di imprevisti (come un file mancante o un CSV corrotto), la fase di caricamento dei dati grezzi è protetta da un blocco `try-catch`. Questo meccanismo permette alla pipeline di isolare l'errore, scartare l'evento problematico e passare direttamente al successivo. Al termine dell'elaborazione, i dati di ciascuna sessione vengono organizzati in una struttura gerarchica `data` e salvati nella cartella `Segmenti_Processati`, utilizzando l'`id` dell'evento come nome del file. Inoltre, il codice genera automaticamente una figura a cinque pannelli, utile per validare visivamente il risultato.

Successivamente, ciascun segnale è stato segmentato sulla base degli intervalli temporali definiti nel protocollo (sezione 3.3), estraendo esclusivamente i campioni compresi tra il timestamp di inizio e di fine di ciascuna sessione. Come ulteriore misura di controllo, durante l'elaborazione del primo evento viene analizzato attentamente il vettore temporale dell'accelerometro (ACC). La pipeline verifica che i dati estratti corrispondano effettivamente alla finestra temporale richiesta per la sessione: se rileva una discrepanza, tipicamente dovuta a un problema con i fusi orari, il sistema genera subito un segnale di allarme.

4.2 Decomposizione dell'EDA

La scomposizione del segnale EDA nelle sue componenti tonica e fasica è stata implementata dalla collega Silvia Barbera nell'ambito del medesimo progetto di tesi. Ne viene qui presentata una sintesi funzionale perché le feature estratte da questa decomposizione costituiscono parte dell'input dei modelli predittivi descritti nel capitolo successivo.

Il segnale EDA è la sovrapposizione di due processi fisiologici distinti: la componente tonica (SCL, *Skin Conductance Level*), un andamento lento legato all'idratazione basale della pelle e alla termoregolazione, e la componente fasica (SCR, *Skin Conductance Response*), i picchi rapidi generati dall'attivazione del sistema nervoso simpatico in risposta a stimoli emotivi o cognitivi. Se non separate, le due componenti si sovrappongono e rendono impossibile distinguere lo stress dall'idratazione cutanea.

Per effettuare questa separazione è stato adottato l'algoritmo cvxEDA, proposto da Greco et al. (2016) [28] e implementato in MATLAB tramite il solver quadprog. A differenza dei classici filtri passa-basso, cvxEDA affronta la decomposizione come un problema inverso: risale dall'EDA osservata all'attività nervosa simpatica che l'ha generata, sfruttando un modello fisiologico della ghiandola sudoripara basato sulla funzione di Bateman, una doppia esponenziale con costante di ascesa $\tau_1 = 0.7$ s e di discesa $\tau_0 = 3.0$ s. La componente tonica è modellata tramite una base di B-spline cubiche con nodi ogni 10 s, sufficientemente lenta da seguire le derive basali senza inseguire i picchi fasici.

La decomposizione è formulata come un problema di ottimizzazione quadratica convessa con due termini di regolarizzazione: una penalità L_1 che impone la sparsità del driver nervoso (solo impulsi discreti e significativi), e una penalità L_2 sulla componente tonica che ne impedisce le variazioni brusche. Il vincolo di non-negatività sul driver riflette la natura esclusivamente eccitatoria del simpatico sulle ghiandole sudoripare.

Dal punto di vista implementativo, il solver utilizzato è `interior-point-convex`; qualora non converga, condizione che si verifica tipicamente con segnali di ampiezza molto ridotta come quelli del soggetto 2, la pipeline ritenta automaticamente con l'algoritmo `trust-region-reflective`, numericamente più conservativo. Se entrambi i tentativi falliscono, la decomposizione viene eseguita tramite un filtro di Butterworth del terzo ordine passa-basso a 0.05 Hz, con la componente fasica ottenuta per sottrazione dalla tonica.

4.3 Elaborazione dell'accelerometro

A partire dai tre assi dell'accelerometro (ACC_x , ACC_y , ACC_z), è stato calcolato il modulo dell'accelerazione risultante come norma euclidea dei tre segnali, normalizzata per la sensibilità nominale del sensore (64 LSB/g). Questo segnale scalare, denominato VMU (*Vector Magnitude Units*), fornisce una stima dell'intensità complessiva del movimento indipendente dall'orientamento del dispositivo sul polso:

$$VMU = \sqrt{ACC_x^2 + ACC_y^2 + ACC_z^2} \quad (4.1)$$

Il VMU è stato utilizzato sia come segnale ausiliario per la caratterizzazione del movimento, sia come indicatore qualitativo per l'identificazione delle finestre temporali maggiormente disturbate da artefatti [3].

4.4 Preparazione del segnale HRV

A partire dalla sequenza degli IBI fornita dal sensore, è stata verificata la disponibilità di un numero sufficiente di battiti validi per il calcolo delle metriche di variabilità

cardiaca. In accordo con le linee guida per l'analisi dell'HRV [16], sono state considerate valide unicamente le finestre contenenti almeno 5 intervalli RR (corrispondenti a 6 battiti consecutivi); in caso contrario, i valori vengono marcati come non disponibili (NaN) ed esclusi dall'analisi successiva.

Le tre metriche HRV calcolate dalla funzione `calc_hrv` sull'intera sessione sono:

- **RMSSD** (*Root Mean Square of Successive Differences*): radice quadrata della media dei quadrati delle differenze tra intervalli RR consecutivi, espressa in millisecondi. È il principale indicatore dell'attività parasimpatica ad alta frequenza [16].
- **SDNN** (*Standard Deviation of NN intervals*): deviazione standard dell'intera serie di intervalli RR in millisecondi, indicatore della variabilità complessiva del ritmo cardiaco.
- **meanNN**: durata media degli intervalli RR in millisecondi, inversamente proporzionale alla frequenza cardiaca media e utilizzata come riferimento per la normalizzazione delle metriche di variabilità.

Il calcolo di queste metriche su finestre temporali scorrevoli, necessario per l'estrazione delle feature, è descritto nel capitolo successivo.

4.5 Qualità del dato finale

Al termine della pipeline, i segnali disponibili per l'estrazione delle feature sono: la componente tonica dell'EDA (SCL), la componente fasica dell'EDA (SCR), il driver fasico (impulsi nervosi sparsi), la frequenza cardiaca istantanea (HR), la sequenza degli IBI validati e il modulo dell'accelerazione (VMU).

La qualità complessiva dei dati varia in modo considerevole tra i soggetti. Il soggetto 3 presenta i tracciati più puliti, con segnali IBI ed EDA entrambi di buona qualità. Il soggetto 2 è penalizzato dai valori basali di conduttanza molto ridotti: la componente fasica ha un'ampiezza che si avvicina al rumore strumentale del sensore, e l'informazione ricavabile dall'EDA risulta di conseguenza limitata. Il soggetto 1 presenta invece instabilità sul segnale accelerometrico causata dagli artefatti vibroacustici, che

ha ridotto il numero di segmenti IBI validi disponibili per il calcolo dell'HRV. Queste differenze sono state tenute in considerazione nella fase di estrazione delle feature, come discusso nel capitolo successivo.

4.6 Analisi qualitativa dei segnali pre-processati

Al termine della pipeline, i segnali elaborati sono stati visualizzati per ciascuna sessione tramite grafici che sovrappongono la componente tonica dell'EDA (SCL), il driver fasico ricostruito da cvxEDA, la frequenza cardiaca (HR), il modulo dell'accelerazione (VMU) e il valore di RMSSD calcolato sull'intera sessione. L'analisi qualitativa di questi grafici ha un duplice scopo: verificare la correttezza della pipeline di pre-processing e caratterizzare il profilo fisiologico individuale di ciascun soggetto in ciascuna condizione, informazione necessaria per la successiva definizione delle regole cliniche del modello predittivo.

4.6.1 Sessioni del 16 novembre 2022

Soggetto 2 – Baseline sensoriale

La sessione documenta un'interazione libera con oggetti luminosi a variazione cromatica. Dopo un iniziale periodo di stabilizzazione tecnica del segnale (16:20–16:25), la componente tonica (SCL) si assesta su un plateau elevato. In linea con il cosiddetto "paradosso dell'oggetto" descritto da Prince et al. [29], questo plateau non indica un sovraccarico sensoriale ma un coinvolgimento attivo con gli stimoli (fenomeno dello *stimming*). La componente fasica mostra picchi frequenti di ampiezza media in corrispondenza della manipolazione di fibre ottiche e palline (16:25, 16:53, 17:02), morfologia tipica dell'esplorazione autonoma [24]. La coincidenza temporale tra i picchi accelerometrici e le risposte elettrodermiche conferma che l'attivazione fisiologica è generata dall'interazione motoria volontaria con gli stimoli, non ad artefatti. La frequenza cardiaca si mantiene su valori medi di 85–90 BPM, compatibili con uno stato di veglia attiva e non di stress. Il valore di RMSSD pari a 53 ms indica una solida regolazione parasimpatica [5]. La presenza contemporanea di un'elevata attività EDA

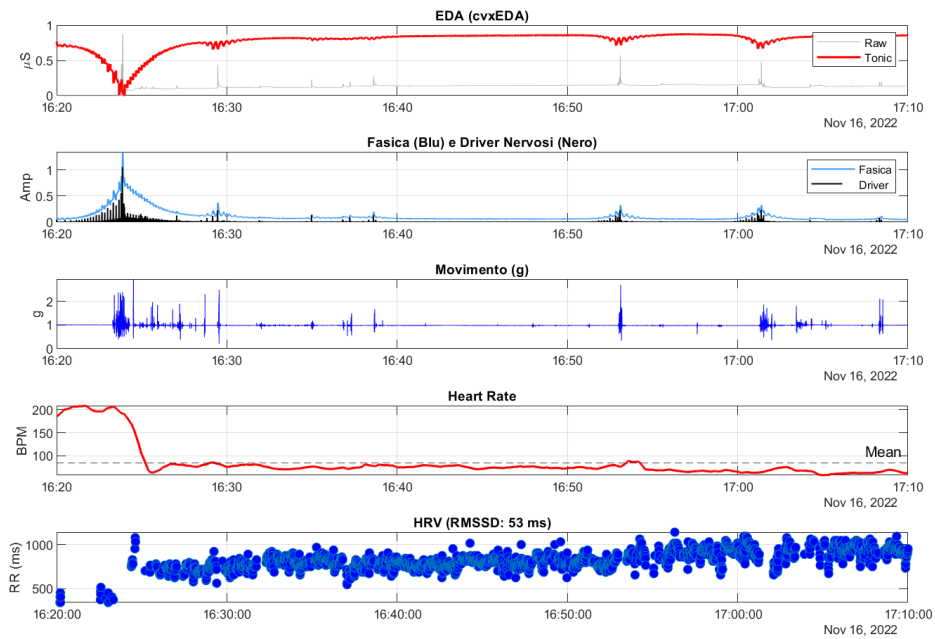


Figura 4.1: Segnali pre-processati del soggetto 2, sessione baseline del 16 novembre 2022. Dal-l'alto: componente tonica dell'EDA (SCL), driver fascico (SCR), frequenza cardiaca (HR) e modulo dell'accelerazione (VMU). Il valore di RMSSD di sessione è 53 ms.

e di una buona HRV suggerisce un quadro fisiologico favorevole: il soggetto riesce a gestire stimoli sensoriali intensi senza perdere il corretto equilibrio del sistema nervoso autonomo.

Soggetto 1 – Sessione sensoriale (social context switch)

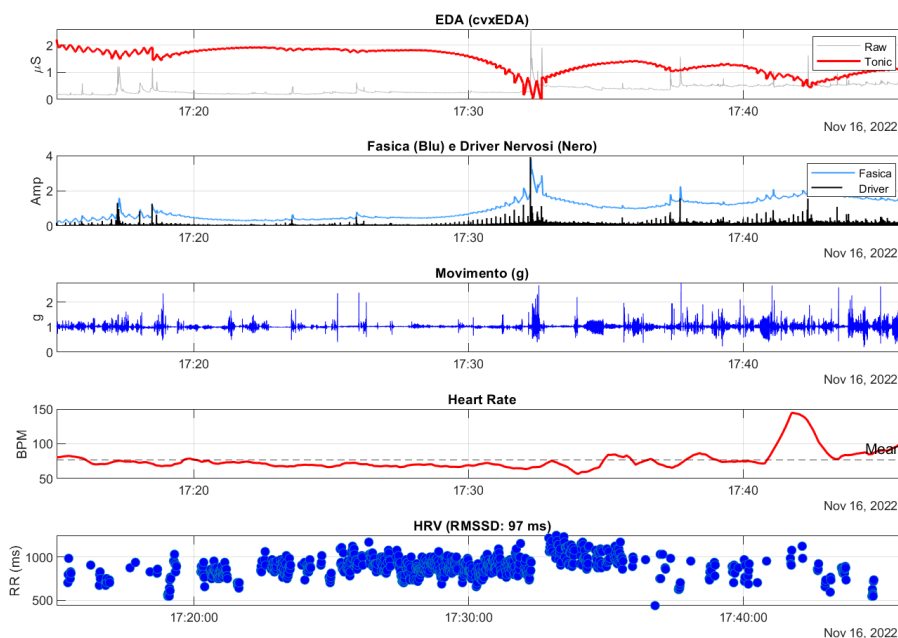


Figura 4.2: Segnali pre-processati del soggetto 1, sessione sensory del 16 novembre 2022. Il valore di RMSSD di sessione è 97 ms.

La sessione presenta due fasi fisiologicamente distinte. Nella fase passiva (17:10–17:32), la componente tonica (SCL) è piatta e stabile, coerentemente con l’efficacia della stimolazione vibroacustica nel mantenere un arousal regolato. Con l’inizio del gioco sociale (17:33) si osserva un’improvvisa e intensa attivazione della componente fasica (SCR): in accordo con Welch (2012) [6], questo picco rappresenta un *social context switch*, ossia un’attivazione del sistema simpatico funzionale alla mobilitazione delle risorse per l’interazione, e non un segnale di sofferenza. Il lieve crollo della componente tonica in prossimità del picco è attribuibile a un artefatto da cambio posturale, correttamente distinto dal segnale biologico sottostante dalla pipeline cvxEDA. Nonostante l’aumento dell’attività motoria e il picco elettrodermico, la frequenza cardiaca rimane stabile su valori medio-bassi: questo disaccoppiamento, ovvero alta attivazione simpatica cutanea associata a frequenza cardiaca contenuta, indica che l’attività motoria ha funzione autoregolativa [4]. Il valore di RMSSD di 97 ms è elevato, indicativo di una regolazione parasimpatica superiore alla media [5].

Soggetto 3 – Sessione recovery (recupero post-farmaco)

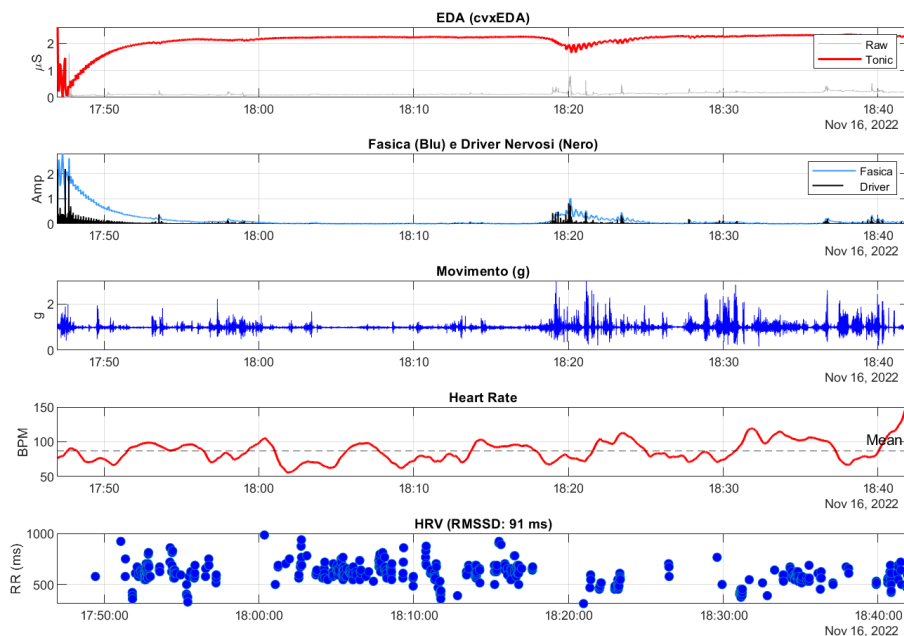


Figura 4.3: Segnali pre-processati del soggetto 3, sessione recovery del 16 novembre 2022. Il valore di RMSSD di sessione è 91 ms. La linea tratteggiata indica l'intervento farmacologico (17:55).

Il grafico mostra il passaggio da una fase iniziale di forte attivazione simpatica, ancora presente dopo l'evento acuto, a uno stato di sedazione profonda indotto dai farmaci. All'inizio (17:48–17:55), sia la componente tonica sia quella fasica dell'EDA restano molto elevate, segno che ormoni come cortisolo e adrenalina sono ancora in circolo [30]. Intorno alle 17:55, però, il segnale crolla rapidamente e l'attività fasica scompare: questo silenzio è tipico dell'effetto delle benzodiazepine, che riducono l'attività dei nervi sudomotori tramite i recettori GABA. Questa fase è particolarmente utile anche dal punto di vista analitico, perché rappresenta un riferimento chiaro di come il segnale decresce quando il sistema simpatico viene spento farmacologicamente.

Durante la sedazione, la frequenza cardiaca varia tra 60 e 100 BPM, indicando che il controllo vagale sta gradualmente tornando attivo. Il picco finale, che arriva intorno ai 150 BPM insieme all'aumento del VMU, mostra che il sistema cardiovascolare non è bloccato, ma mantiene la capacità di reagire [7]. Il valore elevato di RMSSD

(91 ms) conferma un forte recupero dell'equilibrio fisiologico: nel complesso, l'intervento farmacologico ha spostato il soggetto da una condizione più rischiosa (EDA alta e HRV bassa) a una più protettiva (EDA molto bassa e HRV elevata), in accordo con il modello di Fenning et al. [5].

4.6.2 Sessioni del 18 novembre 2022

Soggetto 3 – Sessione work (pressione profonda)

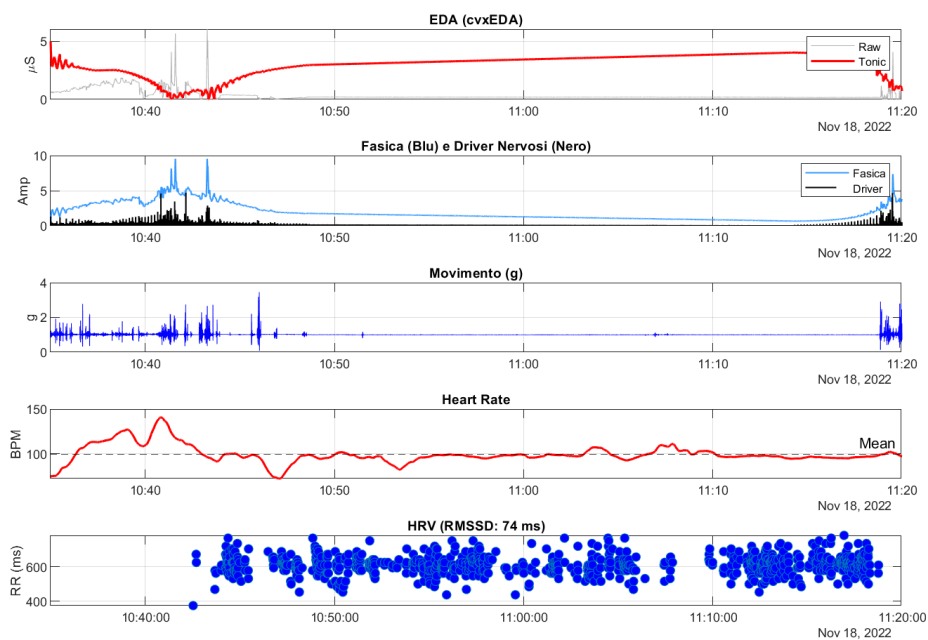


Figura 4.4: Segnali pre-processati del soggetto 3, sessione work del 18 novembre 2022. Il valore di RMSSD di sessione è 74 ms.

Questa sessione evidenzia gli effetti fisiologici della stimolazione somatica profonda. Nella fase preparatoria (10:35–10:45) si osservano diversi picchi dell'EDA associati al movimento rilevato dal VMU, con un artefatto transitorio intorno alle 10:43 che viene correttamente gestito dalla pipeline. Durante la fase centrale, in cui viene applicata la pressione (10:50–11:15), il segnale del VMU diventa quasi completamente piatto e, per oltre 20 minuti, non si rileva alcuna attività fasica.

Questo dato è particolarmente interessante: secondo i modelli di attivazione difensiva [6], il contatto fisico può rappresentare uno stimolo stressante nei soggetti con

ASD. Tuttavia, la totale assenza di attività fasica durante il contatto suggerisce che la stimolazione profonda non induce una risposta di allarme, ma al contrario la riduce in modo efficace. Nel frattempo, la componente tonica dell'EDA rimane su valori sostenuti, con una lieve tendenza all'aumento. La presenza di un livello tonico elevato, insieme all'assenza di risposte fasiche, non indica disagio, ma piuttosto uno stato di attivazione metabolica legata all'attenzione più che alla difesa [30]. Anche la frequenza cardiaca si mantiene su valori di riposo, senza episodi di tachicardia reattiva, mentre l'RMSSD pari a 74 ms indica un buon tono vagale. Nel complesso, emerge un quadro di co-attivazione equilibrata dei due sistemi autonomi: il sistema simpatico supporta l'attenzione, mentre il parasimpatico garantisce la stabilità cardiaca [5]. Al termine della sessione (11:20), la rimozione del sensore coincide con la ripresa del movimento e dell'attività fasica, come previsto alla fine della procedura.

Soggetto 2 – Sessione sensoriale (luci di Wood)

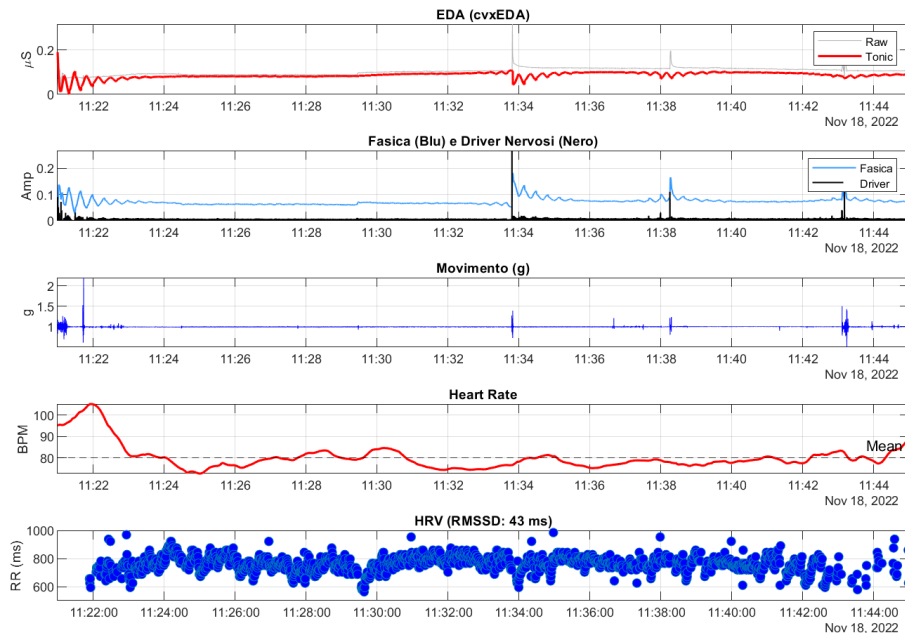


Figura 4.5: Segnali pre-processati del soggetto 2, sessione sensory del 18 novembre 2022. Il valore di RMSSD di sessione è 43 ms.

Questo tracciato rappresenta il caso tecnicamente più complesso dell'intera raccolta dati, relativo a un soggetto con bassa responsività elettrodermica (cfr. sezione 3.4). I valori della componente tonica dell'EDA sono infatti molto bassi, nell'ordine di $0.1 \mu\text{S}$, quindi prossimi al limite del rumore strumentale. Nonostante ciò, l'algoritmo cvxEDA è riuscito a individuare quattro distinti eventi fasici agli orari 11:21, 11:34, 11:38 e 11:43 [28]. È importante distinguere questo profilo da quello di reattività attenuata descritto da Panju et al. [24], spesso associato a stati di chiusura o ansia cronica. In questo caso, infatti, la presenza di picchi fasici ben definiti e sincronizzati con eventi motori indica che la capacità reattiva del sistema nervoso è conservata, anche se con un'ampiezza elettrodermica ridotta. Il segnale VMU risulta per lo più stabile, con variazioni che compaiono in stretta coincidenza con i picchi fasici dell'EDA: questo allineamento suggerisce che l'attività osservata non è dovuta ad artefatti, ma è legata alla manipolazione mirata di oggetti sensoriali. Intorno alle 11:22 si registra un aumento iniziale della frequenza cardiaca (oltre 100 BPM), seguito da una stabilizzazione tra 75 e 85 BPM. La dissociazione tra attività fasica e frequenza cardiaca durante gli episodi di manipolazione (11:34, 11:38) suggerisce un coinvolgimento sensoriale positivo [29]: l'interesse verso l'oggetto genera una risposta EDA, ma l'assenza di tachicardia esclude una reazione di stress. Il valore di RMSSD pari a 43 ms indica un ritiro vagale funzionale, necessario per sostenere l'attenzione focalizzata, piuttosto che un segnale di disagio [5].

Soggetto 1 – Sessione vibroacustica

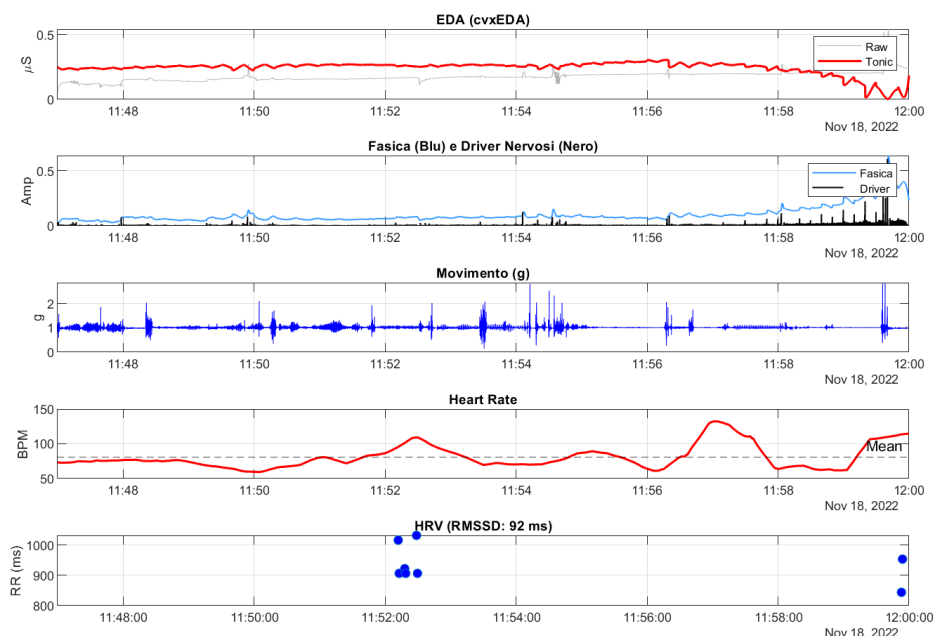


Figura 4.6: Segnali pre-processati del soggetto 1, sessione del 18 novembre 2022. Il valore di RMSSD sui segmenti validi è 92 ms.

Questo tracciato rappresenta la sfida più significativa dell'intera raccolta dati, dovuta alle interferenze meccaniche generate dal letto vibroacustico sui sensori. Il segnale EDA grezzo presenta infatti rapide oscillazioni ad alta frequenza, causata dalla trasmissione diretta delle vibrazioni. Nonostante ciò, l'algoritmo basato su ottimizzazione convessa è riuscito a filtrare efficacemente il rumore, restituendo una componente tonica pulita e stabile per tutta la durata della registrazione (11:47–11:58), intorno a $0.25 \mu\text{S}$ [28]. Questa stabilità riflette uno stato di equilibrio fisiologico: il soggetto non mostra né segnali di stress né una disattivazione marcata.

L'assenza di attività fasica di fondo conferma che le vibrazioni continue non vengono interpretate erroneamente come stimoli nervosi. Si osservano invece due brevi episodi di attivazione fasica (11:50 e 11:53), in corrispondenza dello spostamento della vibrazione verso la regione toracica e della stimolazione a corpo intero. In linea con quanto descritto da Welch [6], questi eventi possono essere interpretati come risposte di orientamento allo stimolo, piuttosto che reazioni difensive: lo suggeriscono

sia l'assenza di un aumento tonico prolungato sia il mantenimento della variabilità cardiaca.

Il segnale VMU registra un'elevata attività meccanica esterna che però non si accompagna a un aumento della frequenza cardiaca. Questo disaccoppiamento tra movimento e risposta cardiovascolare rappresenta una firma tipica della stimolazione passiva ed è particolarmente utile come esempio negativo per il classificatore [4].

Gli aumenti transitori della frequenza cardiaca (11:52, 11:57) coincidono con i momenti di stimolazione più intensa e rientrano rapidamente ai valori basali, indicando una buona capacità di recupero vagale [7]. Eventuali segmenti mancanti del segnale durante le vibrazioni più forti non sono da considerarsi errori di acquisizione, ma una conseguenza dell'intensità dello stimolo; la pipeline li gestisce escludendo gli intervalli IBI non validi (sezione 4.4). Infine, il valore di RMSSD pari a 92 ms, calcolato sui segmenti validi, indica una marcata dominanza vagale: anche in presenza di stimolazione intensa, il sistema parasimpatico mantiene un controllo efficace, riducendo la probabilità di risposte comportamentali disfunzionali [5].

5. Estrazione delle feature

La pipeline di estrazione delle feature è implementata in `feature_extraction.m` e si articola in tre fasi sequenziali: calibrazione della baseline individuale, segmentazione a finestre scorrevoli con estrazione delle feature, e visualizzazione dei risultati estratti. L'output finale è un dataset in formato `.csv` contenente una riga per ogni finestra temporale elaborata, con le feature fisiologiche e l'etichetta associata.

5.1 Calibrazione della baseline individuale

Prima dell'estrazione delle feature, la pipeline esegue una fase di calibrazione per determinare i valori fisiologici di riferimento di ciascun soggetto, necessari al calcolo delle feature. La sessione di riferimento è definita in modo personalizzato per ogni individuo: il soggetto 3 utilizza la sessione di recupero post-farmaco, il soggetto 1 la sessione vibroacustica e il soggetto 2 la sessione di baseline.

Per ciascuna sessione di riferimento vengono calcolati tre valori medi: la frequenza cardiaca di riferimento (`ref.HR`), calcolata escludendo i valori inferiori a 50 BPM per scartare le perdite di contatto del sensore PPG; il livello di conduttanza cutanea di riferimento in scala logaritmica (`ref.EDA_Log`), calcolato come $\log(\overline{SCL} + 1)$ escludendo i campioni inferiori a $0.05 \mu\text{S}$; e la magnitudo media dell'accelerazione di riferimento (`ref.ACC`).

5.2 Segmentazione a finestre scorrevoli

L'estrazione delle feature è condotta su finestre temporali scorrevoli di 60 secondi con overlap di 30 secondi [3], producendo una finestra ogni 30 s per tutta la durata di ciascuna sessione. Nel caso in cui la sessione sia troppo breve per contenere una finestra completa ma superi i 30 secondi, viene comunque estratta una singola finestra a partire dall'inizio.

Prima dell'estrazione, ciascuna finestra viene sottoposta a un controllo di qualità sull'EDA tonica: se il valore medio della componente tonica nella finestra risulta inferiore a $0.05 \mu\text{S}$, la finestra viene scartata e non inclusa nel dataset. Questa soglia è

stata adottata per gestire i segmenti del soggetto 2, evitando di includere finestre in cui il segnale EDA è indistinguibile dal rumore strumentale.

5.3 Feature estratte

Per ogni finestra temporale valida vengono calcolate le seguenti feature, raggruppate per segnale sorgente.

5.3.1 Feature di frequenza cardiaca

Dalla finestra del segnale HR fornito dal sistema di acquisizione del braccialetto Empatica E4 vengono estratte tre feature. La prima è HR_mean, la frequenza cardiaca media nella finestra, calcolata come \overline{HR} sui campioni disponibili. La seconda è HR_std, la deviazione standard della frequenza cardiaca nella finestra, indicatore della variabilità a breve termine. La terza è Delta_HR, lo scostamento dalla baseline individuale, calcolato come $\Delta HR = \overline{HR} - HR_{ref}$.

Motivazione fisiologica per l'ASD. La frequenza cardiaca rappresenta il parametro più diretto dell'attivazione simpatica disponibile tramite PPG non invasivo. Nei soggetti con ASD, l'HR non è solo un indicatore di sforzo fisico ma riflette stati emotivi e comportamentali interni che il soggetto non è in grado di comunicare verbalmente: Groden e colleghi hanno documentato l'elevazione sostenuta di HR durante episodi di agitazione comportamentale in bambini con ASD [7], mentre Lydon e colleghi [4] hanno mostrato che l'HR si eleva significativamente prima dell'inizio di comportamenti problematici, rendendola una feature predittiva e non solo reattiva. HR_std cattura l'instabilità intra-finestra: una maggiore variabilità nell'HR su scala di 60 secondi è stata associata a stati di ipervigilanza, nei quali il sistema nervoso autonomo oscilla rapidamente tra attivazioni simpatiche e parasimpatiche [13]. Delta_HR normalizza le differenze inter-soggetto nel tono cardiaco basale, problema rilevante in questa popolazione dato che i tre soggetti presentano profili di frequenza a riposo molto diversi tra loro.

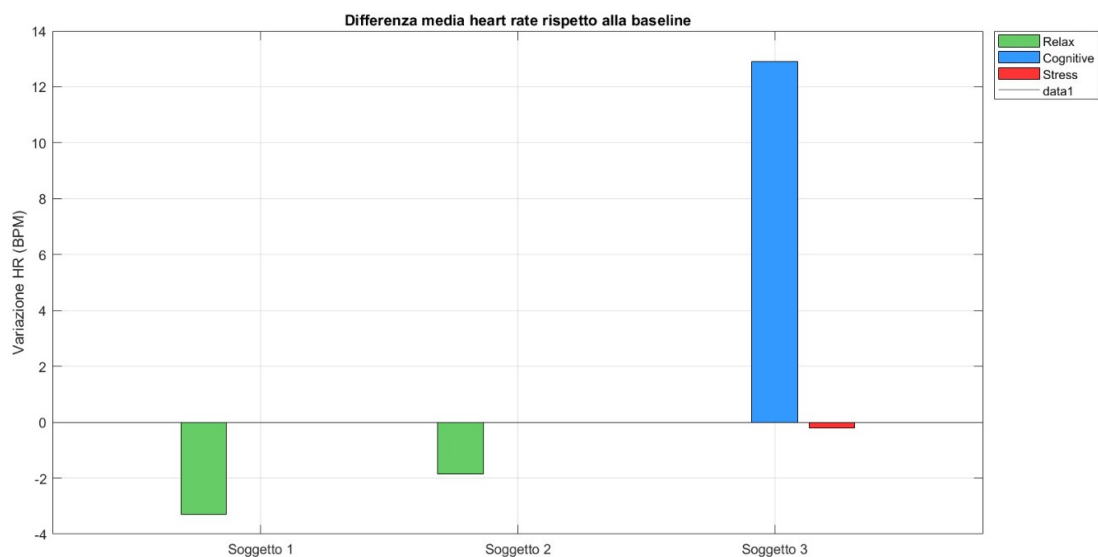


Figura 5.1: Variazione media ΔHR rispetto alla baseline per soggetto e condizione (verde: relax, blu: sovraccarico cognitivo, rosso: stress).

5.3.2 Feature EDA tonica

Le feature della componente tonica sono calcolate in scala logaritmica per comprimere la distribuzione asimmetrica dei valori e rendere comparabili i soggetti con livelli basali molto diversi, condizione documentata in letteratura per i bambini con ASD [24]. EDA_tonic_log rappresenta il livello tonico medio nella finestra in scala logaritmica, $\log(\overline{SCL} + 1)$, mentre Delta_EDA_log misura lo scostamento dalla baseline individuale nella stessa scala, $\Delta EDA = EDA_Tonic_Log - EDA_Log_{ref}$. La terza feature, EDA_slope, è la pendenza della componente tonica nella finestra, stimata tramite regressione lineare del primo ordine sul vettore dei campioni SCL: un valore positivo indica un'attivazione simpatica in corso, un valore negativo un ritorno verso il livello basale [30].

Motivazione fisiologica per l'ASD. La componente tonica dell'EDA (SCL, *Skin Conductance Level*) rappresenta il livello di attivazione simpatica di base delle ghiandole eccrine, che sono controllate esclusivamente dal sistema nervoso simpatico. Nei bambini con ASD, questo parametro risulta spesso alterato. In particolare, Fenning et al. [5] hanno evidenziato che livelli tonici cronicamente elevati sono associati a

una maggiore frequenza di comportamenti esternalizzanti. Allo stesso tempo, Kushki et al. [13] hanno mostrato che la componente tonica può distinguere diversi stati di *arousal* anche quando le risposte fasiche non sono chiaramente visibili.

L'applicazione della trasformazione logaritmica è giustificata dalla presenza dei cosiddetti *low responders*. Panju et al. [24] hanno infatti osservato che una quota significativa di bambini con ASD presenta valori basali di conduttanza molto bassi, con distribuzioni fortemente asimmetriche che rendono poco affidabili i test parametrici tradizionali.

Un esempio concreto è rappresentato dal soggetto 2 del dataset, che mostra valori estremamente ridotti di attività elettrodermica. Nelle sessioni di relax, il valore medio di `EDA_tonic_log` è pari a $0.57 \log(\mu\text{S})$ nella sessione del 16 novembre e scende fino a $0.08 \log(\mu\text{S})$ in quella del 18 novembre, indicando una responsività prossima al limite del rumore strumentale, come già discusso nella sezione 3.4.

Infine, la variabile `EDA_slope` descrive l'andamento nel tempo della componente tonica e fornisce informazioni sulla dinamica dello stato fisiologico. Ferguson et al. [30] hanno mostrato che una pendenza positiva ($\text{slope} > 0$) è spesso associata a un aumento progressivo dell'attivazione, che può precedere un'escalation comportamentale, mentre una pendenza negativa indica una fase di recupero o de-escalation.

5.3.3 Feature EDA fasica e driver

Le feature della componente fasica e del driver fasico quantificano l'entità e la densità delle risposte elettrodermiche nella finestra. `Phasic_max` corrisponde all'ampiezza massima della componente fasica nella finestra, $\max(\text{SCR})$. `Phasic_AUC` è l'area sotto la curva della componente fasica in valore assoluto, calcolata tramite integrazione numerica trapezoidale:

$$\text{AUC} = \int_{t_0}^{t_1} |\text{SCR}(t)| dt$$

e rappresenta il carico emotivo complessivo nella finestra. `Driver_peaks` è infine il numero di campioni del driver fasico con ampiezza superiore a 0.01, utilizzato come stima della densità degli impulsi del sistema nervoso simpatico nella finestra.

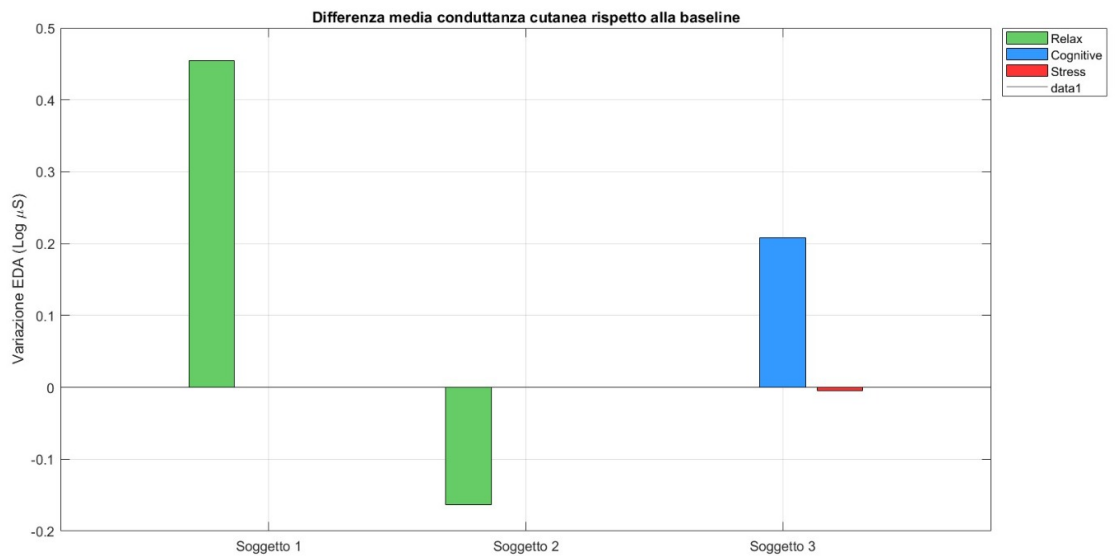


Figura 5.2: Variazione media Δ EDA log rispetto alla baseline per soggetto e condizione (verde: relax, blu: sovraccarico cognitivo, rosso: stress).

Motivazione fisiologica per l'ASD. Le risposte fasiche dell'EDA (*SCR*, *Skin Conductance Response*) rappresentano eventi puntuali di attivazione del sistema simpatico: ciascuna risposta è legata a una scarica del nervo sudomotore verso le ghiandole eccrine, con una latenza di circa 1–3 secondi rispetto allo stimolo che l'ha generata. Nei soggetti con ASD, sia la frequenza sia l'intensità di queste risposte sono state messe in relazione con la gravità dei comportamenti dirompenti [3], ma anche con il livello di coinvolgimento durante attività terapeutiche strutturate [14].

La metrica *Phasic_AUC* consente di integrare il carico emotivo complessivo all'interno di una finestra temporale e risulta più stabile rispetto a singole misure di picco, soprattutto quando il segnale è rumoroso [14].

Infine, *Driver_peaks*, derivato dal modello *cvxEDA* [28], fornisce una stima diretta della frequenza degli impulsi nervosi sudomotori. Questo indicatore è particolarmente utile perché rimane robusto anche quando la decomposizione del segnale non è perfetta, come spesso accade nei dati acquisiti al polso.

5.3.4 Feature di movimento

Dal segnale accelerometrico viene estratta una singola feature, *ACC_mean*, che rappresenta la media del modulo VMU all'interno della finestra ed è utilizzata come indicatore dell'intensità media del movimento.

L'accelerometro non viene incluso tanto per misurare direttamente l'attività motoria, quanto piuttosto come riferimento per interpretare correttamente le feature fisiologiche estratte nello stesso intervallo temporale. Come evidenziato da Lydon et al. [4], nei soggetti con ASD è frequente osservare un disaccoppiamento tra frequenza cardiaca e movimento (*HR-ACC decoupling*): durante episodi di disregolazione comportamentale, l'aumento della frequenza cardiaca può essere dovuto al movimento (ad esempio agitazione o stereotipie) più che a un reale aumento dell'arousal autonomico.

In questo contesto, un valore elevato di *ACC_mean* associato a un aumento di *Delta_HR* suggerisce che la variazione cardiaca sia principalmente di origine motoria. Al contrario, un aumento della frequenza cardiaca in presenza di basso movimento indica più probabilmente una risposta autonoma. Questa integrazione tra feature consente al classificatore di distinguere le due componenti senza la necessità di applicare filtri preliminari al segnale.

5.3.5 Variabilità della frequenza cardiaca

La feature *HRV_RMSSD* è la radice quadrata della media dei quadrati delle differenze tra intervalli RR consecutivi, espressa in millisecondi. Poiché il calcolo dell'*RMSSD* su finestre di 60 secondi risulta statisticamente instabile, questa feature viene assegnata a ogni finestra come costante di sessione, riportando il valore calcolato sull'intera durata della sessione nella fase di preprocessing (sezione 4.4). Il valore riflette il tono vagale del soggetto nella sessione [5]. L'*RMSSD* è il marcatore temporale del tono parasimpatico più robusto su finestre brevi: riflette la capacità del PNS di modulare rapidamente la frequenza cardiaca in risposta alle variazioni respiratorie (aritmia sinusale respiratoria), che è mediata principalmente dall'attività del nervo vago [17]. Nei soggetti con ASD, Fenning e colleghi [5] riportano valori di *RMSSD* significativamente ridotti rispetto ai controlli, indicativi di una compromissione del *vagal brake*,

ovvero il meccanismo descritto nella teoria polivagale di Porges [9] per cui il PNS sopprime attivamente il SNS nelle situazioni percepite come sicure. La scelta di assegnare l'RMSSD a livello di sessione, anziché calcolarlo per finestra, riflette un compromesso metodologico necessario: su 60 secondi, il numero tipico di intervalli RR disponibili (40–80) è sufficiente per l'RMSSD ma non per gli indici frequenziali LF/HF, che richiedono almeno 256 campioni [16]. Il valore di sessione viene quindi trattato come una covariata fisiologica stabile del soggetto in quella condizione, non come feature dinamica finestra per finestra.

5.4 Dataset finale

Al termine dell'elaborazione, le feature di tutte le finestre valide di tutti i soggetti vengono raccolte in una tabella e salvate nel file `dataset_features_complete.csv`. Ogni riga del dataset corrisponde a una finestra temporale ed è associata al soggetto (`subject`), alla sessione (`session`), al tempo di inizio della finestra (`time`) e all'etichetta di condizione (`label`), con valori 0, 1 e 2 corrispondenti rispettivamente alle condizioni di relax, cognitivo e stress.

Le feature complessivamente presenti nel dataset, con il segnale sorgente e le rispettive unità di misura, sono riassunte nella tabella 5.1.

Tabella 5.1: Feature estratte per ciascuna finestra temporale.

Feature	Segnale	Unità
HR_mean	HR	BPM
HR_std	HR	BPM
Delta_HR	HR	BPM
EDA_tonic_log	SCL	$\log(\mu\text{S})$
Delta_EDA_log	SCL	$\log(\mu\text{S})$
EDA_slope	SCL	$\mu\text{S}/\text{s}$
Phasic_max	SCR	μS
Phasic_AUC	SCR	$\mu\text{S}\cdot\text{s}$
Driver_peaks	Driver	conteggio
ACC_mean	VMU	g
HRV_RMSSD	IBI	ms

5.5 Distribuzione statistica delle feature nel dataset

Il dataset finale comprende 430 finestre temporali valide distribuite su sei sessioni e tre soggetti. La distribuzione tra le condizioni è sbilanciata: 232 finestre (53.9%) appartengono alla classe 0 (relax), 109 (25.3%) alla classe 2 (stress) e 89 (20.7%) alla classe 1 (cognitive load). Questo squilibrio, inevitabile vista la natura osservazionale del protocollo, influenza direttamente la scelta delle metriche con cui valutare il classificatore e rende necessarie tecniche di bilanciamento dei dati, che verranno discusse nel capitolo 6.

5.5.1 Analisi per feature di frequenza cardiaca

Il segnale più discriminante è delta_HR per il soggetto 3: la condizione cognitive (sessione work) produce uno scostamento medio di +12.9 bpm rispetto alla baseline di recovery, mentre la stessa sessione di recovery (stress) mostra un ΔHR quasi nullo (-0.2 bpm), un risultato coerente con la progressiva normalizzazione dell'HR post-farmaco documentata nel capitolo 3. Il soggetto 1 e il soggetto 2 mostrano valori di ΔHR negativi nelle sessioni Relax (-4.4 e -4.9 BPM rispettivamente), indicando che le sessioni di riferimento scelte (vibroacustica e baseline) sovrastimano leggermente l'HR di riposo in quelle condizioni.

5.5.2 Analisi per feature EDA tonica

Il soggetto 3 presenta i valori più elevati di EDA tonica ($EDA_tonic_log \approx 1.13-1.34$), mentre il soggetto 2 nella sessione sensory del 18 novembre mostra una media di soli $0.08 \log(\mu S)$, confermando il profilo di *low responder* descritto nella sezione 3.4. La feature ΔEDA_log comprime efficacemente questa variabilità basale: il soggetto 3 in cognitive mostra uno scostamento di +0.21, nettamente separato dallo stress (-0.005), suggerendo che la stimolazione somatica guidata provochi un'attivazione simpatica duratura maggiore rispetto alla fase post-crisi. Il soggetto 1 in sensory mo-

stra un $\Delta EDA = +0.64$, anomalo rispetto alle altre sessioni Relax, probabilmente riconducibile agli artefatti vibroacustici sul segnale EDA discussi nella sezione 3.4.

5.5.3 Analisi per feature EDA fasica e driver

La Phasic_AUC è la feature con la varianza più elevata nel dataset: il soggetto 3 in cognitive presenta una media di $112.4 \mu S \cdot s$ per finestra, un valore circa 13 volte superiore alla sessione stress dello stesso soggetto ($8.9 \mu S \cdot s$) e circa 20 volte superiore al soggetto 2 in baseline ($6.8 \mu S \cdot s$). Questa dispersione estrema richiederà una normalizzazione robusta nella fase di classificazione. Il Driver_peaks del soggetto 3 in stress supera 200 impulsi per finestra nelle fasi iniziali della sessione (picchi a 228/finestra), diminuendo progressivamente verso 0 con l'effetto del farmaco. Questo andamento temporale che sarà analizzato nella sezione 6.3.

5.5.4 Analisi per feature ACC e HRV

ACC_mean presenta una distribuzione molto compressa attorno a $1.0 g$ per tutti i soggetti, con varianza trascurabile: questo conferma che le sessioni analizzate non contengono episodi di movimento intenso prolungato, e che il segnale accelerometrico svolge in questo dataset un ruolo prevalentemente diagnostico (controllo qualità) piuttosto che discriminativo. Per HRV_RMSSD, il soggetto 1 mostra i valori più elevati (92–97 ms), coerente con un profilo vagale dominante durante la stimolazione sensoriale passiva; il soggetto 2 presenta i valori più bassi (43–52 ms), suggerendo un tono parasimpatico ridotto, che potrebbe contribuire al basso livello di responsività EDA osservato nella stessa popolazione [5].

5.5.5 Considerazioni sulla separabilità del dataset

Dall'analisi delle distribuzioni emergono tre aspetti rilevanti per la fase di classificazione. In primo luogo, le feature differenziali (Delta_HR, Delta_EDA_log) risultano più informative rispetto alle feature assolute, confermando l'utilità della normalizzazione rispetto alla baseline individuale. In secondo luogo, il soggetto 3 rappresenta la parte più informativa del dataset (198 finestre, con presenza di tutte e tre le classi), mentre il

soggetto 1 contribuisce con sole 86 finestre appartenenti esclusivamente alla classe 0, limitando di fatto la capacità di generalizzazione inter-soggetto.

Infine, si osserva che le feature EDA del soggetto 2 nella sessione sensory del 18 novembre presentano valori anomali ($EDA_tonic_log < 0.1$, prossimi alla soglia di esclusione $min_EDA = 0.05$). Questo suggerisce che il filtro di qualità dell'EDA ha mantenuto alcune finestre al limite dell'affidabilità strumentale, con il rischio di introdurre rumore nella fase di addestramento del classificatore.

6. Modelli predittivi

Il dataset acquisito e descritto nei capitoli precedenti non contiene episodi conclamati di aggressività. L'unico evento clinico documentato è l'attacco di panico del soggetto 3, avvenuto nei minuti precedenti l'inizio della registrazione della sessione *recovery*, della quale sono disponibili solo i dati successivi alla somministrazione del farmaco sedativo. Questa condizione, comune nella ricerca psicofisiologica su soggetti con ASD a basso funzionamento, rende impossibile addestrare un classificatore supervisionato su eventi reali. Per affrontare il problema da due direzioni complementari, sono stati sviluppati parallelamente due modelli: un algoritmo non supervisionato basato su baseline dinamica e Change Point Analysis, che non richiede etichette e si adatta in tempo reale al soggetto, e un classificatore Random Forest soggetto-specifico, sviluppato dalla collega Silvia Barbera, che apprende dai momenti di massimo sforzo autonomico disponibili nel dataset. I due modelli, descritti nelle sezioni seguenti, costituiscono la base per l'architettura ibrida presentata nel capitolo successivo.

6.1 Modello non supervisionato: Anomaly Detection

6.1.1 Motivazione della scelta

Un problema fondamentale nel monitoraggio dei soggetti ASD è che lo stato fisiologico di riferimento non è definibile a priori. Come documentato da Kushki et al. (2015) [13], questi soggetti vivono spesso in uno stato di stress cronico e presentano il fenomeno del *covert arousal*: un sovraccarico emotivo interno elevato non percepibile dall'esterno. Adottare una soglia di allarme fissa porterebbe a falsi allarmi continui, poiché il livello di base cambia continuamente durante la giornata e varia da soggetto a soggetto. Per lo stesso motivo, un approccio supervisionato che richieda a un operatore di annotare manualmente i momenti di baseline e di stress non è applicabile in modo affidabile in questa popolazione.

La soluzione adottata si basa sulla stessa logica proposta da Kushki et al. (2015) [13]: definire una baseline dinamica che insegue la normalità del paziente

istante per istante, e rilevare le deviazioni strutturali da essa piuttosto che il superamento di soglie assolute. Questo principio si applica in modo complementare sia all'EDA sia all'HR: mentre la prima riflette l'attivazione simpatica istantanea, la seconda integra l'effetto combinato di eccitazione simpatica e ritiro parasimpatico su una scala temporale leggermente più lenta [17], rendendo i due segnali non ridondanti dal punto di vista informativo. Per identificare il momento esatto in cui questa deviazione diventa statisticamente irreversibile è stata adottata la Change Point Analysis, ispirata al lavoro di Rosenfield et al. (2010) [15], che ha dimostrato come gli eventi acuti non siano preceduti da una lenta salita, ma da una rottura improvvisa dell'equilibrio fisiologico.

6.1.2 Architettura dell'algoritmo

Il modello prende in ingresso il dataset di feature estratto nel capitolo precedente e analizza, per ciascuna sessione, tre segnali: `Delta_EDA_Log`, `HR_Mean` e `HRV_RMSSD`. `Delta_EDA_Log` è l'indicatore primario dell'attivazione simpatica; la trasformazione logaritmica applicata nella fase di estrazione consente di rilevare le micro-fluttuazioni dello stress anche nei soggetti con conduttanza basale molto ridotta. `HR_Mean` traccia la dinamica cardiaca media nella finestra temporale corrente: a differenza dell'EDA, che risponde pressoché istantaneamente all'eccitazione simpatica, l'HR è modulata sia dal sistema simpatico sia da quello parasimpatico [16]. `HRV_RMSSD` viene trattata come costante di sessione: calcolata sull'intera registrazione, essa funge da indicatore tonico del tono vagale basale del soggetto. Un valore elevato riflette una buona regolazione parasimpatica, capace di modulare e attenuare la risposta simpatica agli stimoli; al contrario, un valore ridotto segnala una condizione di vulnerabilità in cui anche perturbazioni di bassa intensità possono favorire l'insorgenza di episodi di disregolazione [9]. L'asse temporale è ricostruito moltiplicando l'indice di finestra per 0.5, ottenendo valori in minuti compatibili con la frequenza di campionamento di 30 secondi per finestra.

Il primo passaggio consiste nel calcolo della baseline dinamica tramite media mobile su una finestra scorrevole di 5 minuti, corrispondenti a 10 campioni. Questa operazione filtra le micro-fluttuazioni fisiologiche naturali, come respiri profondi o micro-movimenti, estraendo il vero andamento di fondo del segnale e aggiornandolo minuto

per minuto. La stessa procedura viene applicata in parallelo ai segnali `Delta_EDA_Log` e `HR_Mean`, in modo che la baseline di ciascun segnale evolva in modo indipendente. Il secondo passaggio applica la Change Point Analysis, tramite la funzione MATLAB `findchangepts`, ai segnali EDA e HR. Il parametro statistico è configurato sulla media del segnale, cosicché l’algoritmo individui i momenti in cui il livello medio subisce una variazione strutturale stabile. Il numero massimo di change point per segnale per sessione è fissato a 2, una scelta che bilancia la sensibilità al rilevamento con la necessità di contenere i falsi positivi. Ogni volta che viene identificata una discontinuità persistente rispetto all’andamento dei minuti precedenti, l’algoritmo colloca un marcatore temporale denominato change point: è questo marcatore, e non il superamento di una soglia assoluta, a definire il precursore dell’evento critico. L’analisi congiunta dei change point su EDA e HR, e in particolare il loro sfasamento temporale relativo, è il meccanismo attraverso cui il modello distingue una risposta di allerta reale da un artefatto da movimento: un picco isolato sull’EDA senza corrispondente variazione dell’HR nei 30–60 secondi successivi viene interpretato come rumore. È opportuno precisare che lo sfasamento tra i segnali osservato nei risultati, che può arrivare a superare il quarto d’ora, non contraddice la latenza fisiologica del sistema simpatico cardiaco, stimabile in poche decine di secondi [17]. La Change Point Analysis non individua l’esatto momento in cui il segnale cambia, ma il momento in cui quel cambiamento risulta stabile e confermato dall’andamento successivo: questo richiede tempo, aggiungendo un ritardo alla latenza fisiologica. Lo sfasamento inter-segnale è quindi la somma di questi due contributi e contiene in sé informazione sulla durata della transizione autonoma.

6.1.3 Risultati e validazione clinica

In assenza di crisi aggressive nel dataset, la validazione è stata condotta come *proof of concept* confrontando i change point rilevati dall’algoritmo con le annotazioni cliniche e i segnali grezzi già analizzati nella fase di preprocessing. Per ogni sessione sono riportati il valore medio di RMSSD (che contestualizza il profilo vagale di fondo) e la sequenza temporale dei change point su EDA e HR, evidenziandone gli sfasamenti significativi.

Soggetto 3 — sessione work (pressione profonda). Questa sessione illustra l'effetto della terapia a pressione profonda: il soggetto è fisicamente fermo, vigile e recettivo agli stimoli dell'operatore. I grafici mostrano una forte stabilità fisiologica complessiva. Il valore medio di RMSSD si attesta su 74 ms, confermando che il sistema cardiovascolare è in uno stato di rilassamento attivo, con il tono vagale pienamente operativo come freno alla reattività simpatica. Le baseline di EDA e HR partono su valori elevati nei primissimi minuti della sessione, per poi decrescere progressivamente e stabilizzarsi in un plateau di bassa attivazione che si mantiene per il resto della seduta, configurando quello che si potrebbe descrivere come un silenzio dinamico del sistema autonomo. Coerentemente con questa dinamica, l'algoritmo concentra tutti i change point nella fase iniziale: vengono rilevate due discontinuità sul segnale HR ai minuti 2 e 7.5, e una sull'EDA al minuto 9. Nei successivi 35 minuti non viene rilevata alcuna ulteriore variazione strutturale. Il change point sull'EDA al minuto 9 marca con precisione la conclusione della fase caotica di arrivo e l'inizio della stimolazione vera e propria. L'assenza di segnalazioni per tutta la durata della terapia rappresenta una prova di affidabilità particolarmente significativa: il sistema riconosce che il contatto fisico dell'operatore non costituisce uno stimolo sociale allarmante [6], ma un input con funzione calmante, senza confonderlo con un precursore di aggressività. La stabilità della HR durante l'intera sessione è inoltre coerente con l'immobilità corporea del soggetto, escludendo che lo stato di quiete osservato sia riconducibile a una risposta di congelamento da paura piuttosto che a un rilassamento genuino [4].

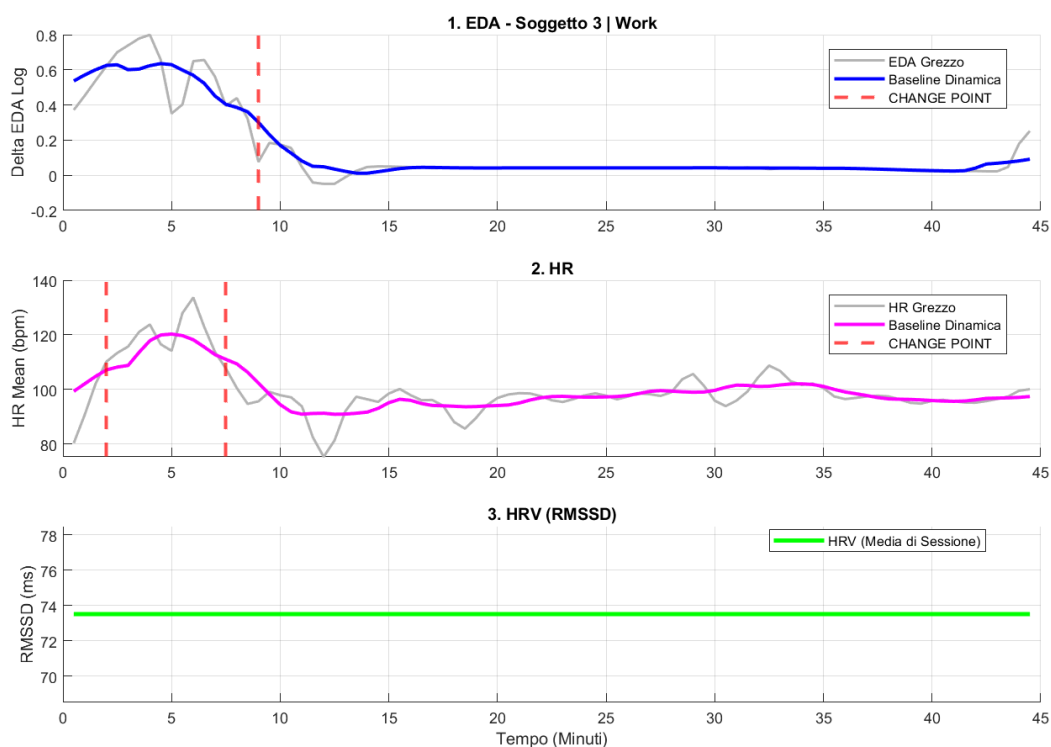


Figura 6.1: Change Point Analysis – sessione S3_soggetto3_work (pressione profonda). Dall’alto: Delta EDA log con baseline dinamica (blu) e change point (rosso tratteggiato) al minuto 9; HR mean con baseline dinamica (magenta) e change point ai minuti 2 e 7.5; HRV RMSSD costante di sessione a 74 ms.

Soggetto 1 — sessione vibro (rumore meccanico). Il paziente è stato sottoposto a terapia su un letto vibroacustico, una condizione che ha introdotto rumore meccanico sui sensori durante l’intera sessione. Nonostante la perdita parziale di dati causata dalle vibrazioni, il valore medio di RMSSD calcolato sui segmenti validi si attesta a 92 ms, confermando una dominanza parasimpatica stabile che rappresenta un fattore protettivo rispetto all’insorgenza di attivazione simpatica acuta.

Il segnale grezzo dell’EDA risulta pesantemente disturbato dal movimento continuo, ma la baseline dinamica si è dimostrata un filtro efficace: la curva restituita è sufficientemente regolare da consentire all’algoritmo di estrarre due change point, localizzati rispettivamente al minuto 2.5 e al minuto 8. Sul segnale HR le discontinuità strutturali emergono più tardi, ai minuti 10 e 11. La sovrapposizione con i segnali grezzi conferma che l’algoritmo ha isolato le uniche variazioni clinicamente rilevanti in un contesto caratterizzato da forte rumore meccanico.

Il change point sull’EDA al minuto 8 coincide con il momento in cui l’operatore

ha spostato la stimolazione vibroacustica verso il torace e i piedi del paziente, intercettando la risposta fisiologica allo stimolo prima che questa raggiungesse un'intensità compatibile con una percezione di minaccia. I change point sull'HR ai minuti 10 e 11 si allineano a un transitorio aumento della frequenza cardiaca seguito da un rapido recupero vagale, escludendo la presenza di un accumulo progressivo di tensione. Si osserva inoltre uno sfasamento di circa due minuti tra l'attivazione simpatica, rilevata sull'EDA al minuto 8, e la risposta cardiovascolare compensatoria, rilevata sull'HR al minuto 10, coerentemente con le dinamiche temporali attese del sistema nervoso autonomo. Di fronte a uno stato di elevata energia meccanica esterna e bassa attivazione cardiaca interna, il modello non ha generato falsi positivi, dimostrando di saper distinguere la vibrazione ambientale dall'agitazione psicomotoria autogenerata.

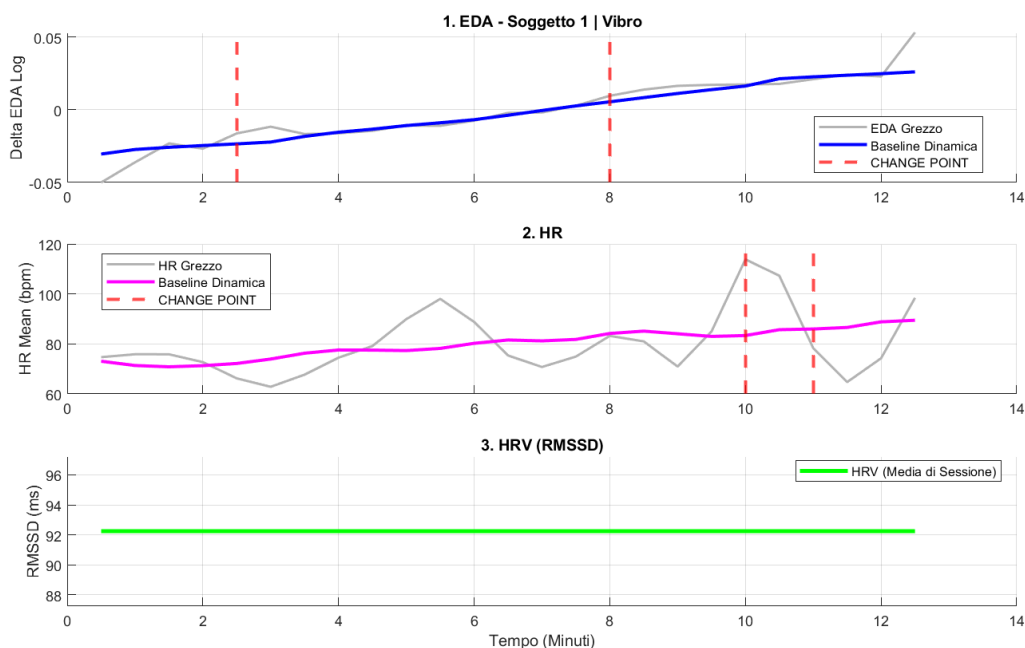


Figura 6.2: Change Point Analysis – sessione S3_soggetto1_vibroacustica (terapia vibroacustica). Dall'alto: Delta EDA log con change point ai minuti 2.5 e 8; HR mean con change point ai minuti 10 e 11; HRV RMSSD costante di sessione a 92 ms.

Soggetto 2 — sessione sensory (low responder). Questa sessione ha messo alla prova l'algoritmo in una condizione limite: la paziente si presenta come soggetto *low responder*, con una conduttanza cutanea basale di ampiezza quasi nulla, durante un compito di attenzione visiva focalizzata. Il quadro complessivo è coerente con uno stato di attenzione sostenuta. Il valore medio di RMSSD è pari a 43 ms, un leggero

ritiro del tono vagale fisiologicamente atteso durante l'impegno cognitivo, che non raggiunge tuttavia livelli compatibili con una condizione di stress.

Sul segnale HR si osserva un picco iniziale di assestamento, intercettato da un change point al minuto 2, seguito da una stabilizzazione attorno agli 80 bpm per il resto della sessione. Sul segnale EDA, nonostante l'ampiezza quasi nulla del segnale grezzo, la baseline dinamica ha restituito una curva in lenta ma costante salita, sulla quale l'algoritmo ha individuato due change point ravvicinati ai minuti 9 e 12.5. La sovrapposizione con i segnali grezzi conferma che il modello riesce a gestire correttamente micro-variazioni nell'ordine di $0.1 \mu\text{S}$.

Il change point iniziale sull'HR isola il tempo di assestamento dei sensori, depurando l'analisi dall'artefatto di inizio sessione. La discontinuità sull'EDA al minuto 12.5 coincide con il principale evento di interazione motoria della sessione: questo suggerisce che la bassa conduttanza basale non riflette un blocco reattivo (*blunted reactivity* [24]), ma che il sistema nervoso autonomo conserva una reattività funzionale agli stimoli rilevanti. L'assenza di tachicardia conferma che la salita dell'EDA rappresenta un fisiologico *engagement* sensoriale e non un accumulo di tensione pre-aggressiva.

Questi dati, pur in assenza di crisi, offrono una proof of concept concreta: la baseline dinamica riesce a estrarre informazione fisiologicamente coerente anche in soggetti con reattività elettrodermica contenuta, confermando la robustezza del metodo in condizioni limite.

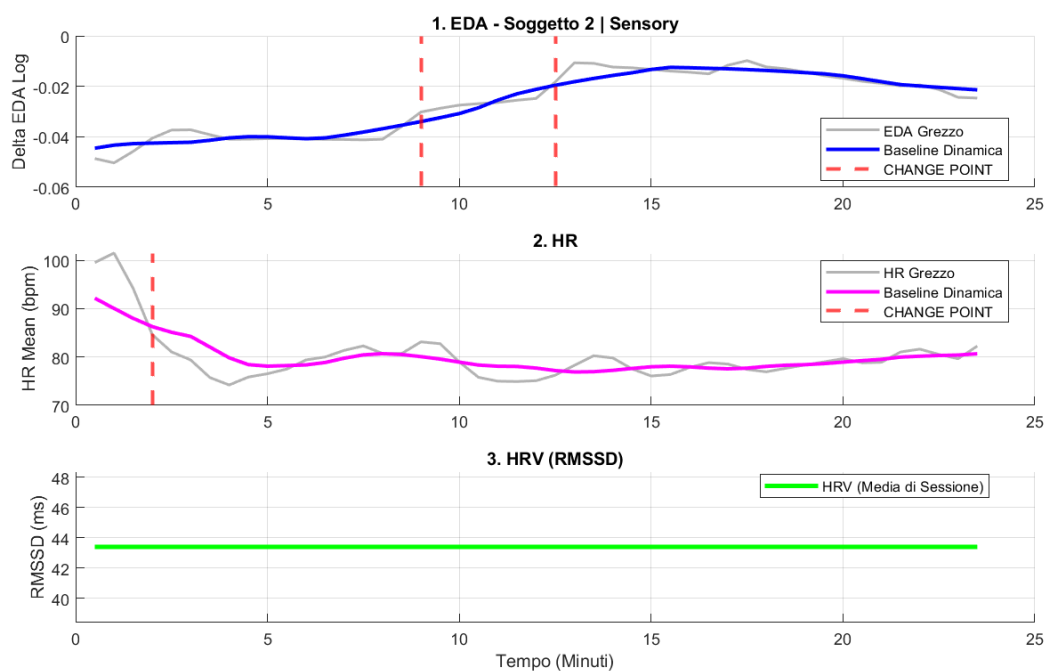


Figura 6.3: Change Point Analysis – Sessione S3_soggetto2_sensory (low responder, attenzione visiva)

- . Dall’alto: Delta EDA log con change point ai minuti 9 e 12.5; HR mean con unico change point al minuto 2 (settling time); HRV RMSSD costante di sessione a 43 ms.

Soggetto 3 – sessione Recovery (effetto farmacologico). Nel dataset clinico disponibile non erano presenti le registrazioni dei minuti antecedenti alla crisi del soggetto, ma unicamente i dati della fase di recupero sotto sedazione farmacologica (Tavor). Questi dati sono stati comunque utilizzati come *proof of concept* rigorosa: un sistema basato sui principi di Kushki e Rosenfield dovrebbe essere in grado di rilevare autonomamente l’imponente salto di regime indotto dal farmaco, senza alcuna informazione a priori sulla sua occorrenza.

Come si evince dai grafici, l’algoritmo ha individuato il change point corrispondente all’inibizione simpatica sul segnale EDA e il successivo risveglio cardiovascolare sul segnale HR. Il valore di RMSSD, indicativo di una dominanza vagale marcata, è coerente con lo stato di sedazione profonda e contestualizza ulteriormente le discontinuità rilevate. Il sistema si è quindi dimostrato sensibile alle rotture strutturali brusche del sistema nervoso autonomo anche in condizioni farmacologicamente indotte.

Questo risultato ha una ricaduta diretta sulla valutazione del modello: la letteratura indica che l'escalation aggressiva produce esattamente questo tipo di discontinuità statistica nel segnale autonomico, in un intervallo temporale antecedente alla manifestazione motoria della crisi [15][13]. Il modello implementato e validato su questi dati è pertanto strutturalmente idoneo a operare come sistema predittivo su dataset che includano la fase pre-criasi.

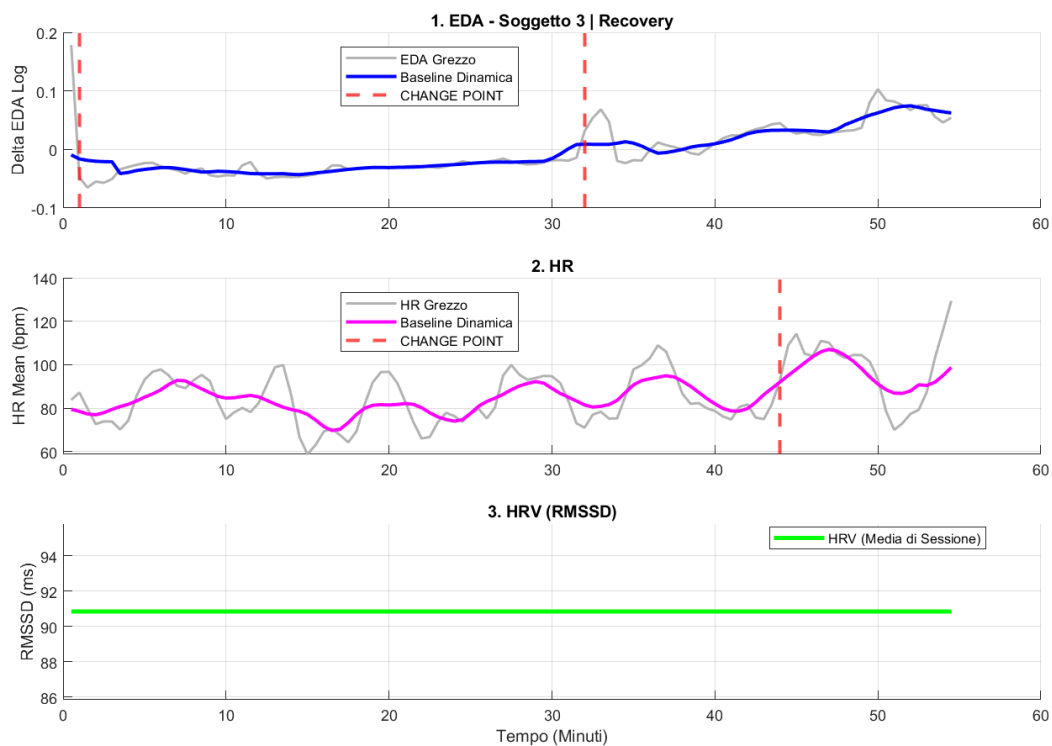


Figura 6.4: Change Point Analysis – Sessione S1_soggetto3_recovery (recupero post-attacco di panico sotto sedazione farmacologica). Dall'alto: Delta EDA log con change point al minuto 1 (inibizione simpatica acuta) e al minuto 32 (transizione verso il recupero); HR Mean con change point al minuto 45 (risveglio cardiovascolare); HRV RMSSD costante di sessione a 91 ms, indicativa di dominanza vagale massiva indotta dal sedativo.

Soggetto 2 — sessione baseline (profilo protettivo). Questa sessione rappresenta una baseline attiva: la paziente è impegnata in un'attività sensoriale stimolante con oggetti luminosi, in uno stato fisiologico coerente con quello che si potrebbe definire stress positivo. Il valore medio di RMSSD si attesta su 53 ms, indicando una regolazione vagale sufficientemente solida da mantenere il sistema autonomo in equilibrio nonostante il coinvolgimento sensoriale. La baseline dinamica dell'EDA mostra una salita iniziale che si stabilizza su un plateau elevato, mantenuto per buona parte

della sessione. Sul segnale HR, il primo change point al minuto 1.5 isola automaticamente il tempo di assestamento iniziale dei sensori. Il change point sull'EDA al minuto 13.5 coincide con l'ingresso nel plateau di massima concentrazione, mentre le discontinuità successive ai minuti 31.5 e 38.5 si allineano al momento in cui la paziente cambia oggetto di interazione, passando alle fibre ottiche. La sovrapposizione con i segnali grezzi conferma la coerenza clinica di questi rilevamenti. In assenza nel dataset dei minuti antecedenti a un vero episodio di crisi, questa sessione è stata utilizzata per testare la specificità del modello. L'algoritmo ha dimostrato di saper isolare i cambiamenti di regime associati all'interazione cognitiva, senza interpretarli come precursori di un accumulo di tensione pre-aggressiva. Questo risultato è rilevante perché conferma che il sistema non si limita a rilevare qualsiasi variazione strutturale del segnale, ma è in grado di discriminare tra attivazioni fisiologicamente attese nel contesto dell'engagement sensoriale e pattern potenzialmente critici.

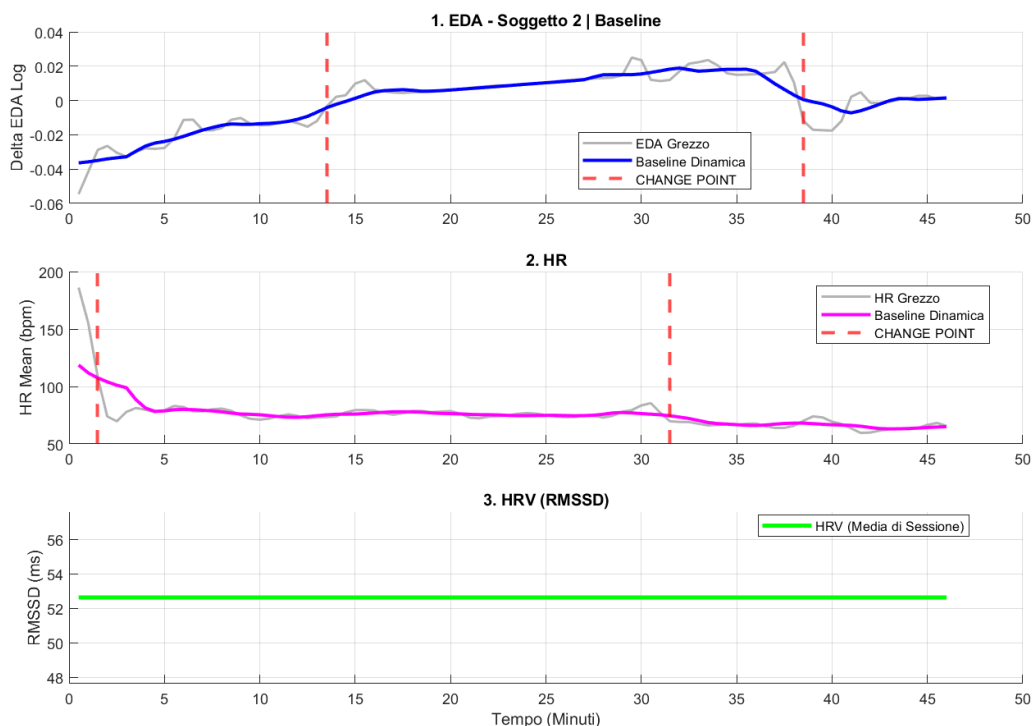


Figura 6.5: Change Point Analysis – Sessione S1_soggetto2_baseline (attività sensoriale stimolante, profilo protettivo). Dall'alto: Delta EDA log con change point ai minuti 13.5 e 38.5; HR mean con change point ai minuti 1.5 (settling time) e 31.5 (cambio di interazione); HRV RMSSD costante di sessione a 52 ms. L'assenza di concordanza temporale tra change point EDA e HR classifica correttamente la sessione come impegno cognitivo sano.

Soggetto 1 — sessione sensory (social context switch). Questa sessione documenta una transizione tra due condizioni distinte: una fase iniziale di profondo rilassamento indotto dalla terapia su letto vibroacustico, seguita da un'interazione sociale attiva. Il valore medio di RMSSD si attesta su 97 ms, indicando una dominanza parasimpatica particolarmente marcata per l'intera durata della sessione. La baseline dinamica dell'EDA si mantiene stabile nella prima fase, per poi mostrare una netta salita nella seconda metà della sessione. L'algoritmo intercetta questa transizione posizionando due change point sull'EDA ai minuti 17 e 22.5. Il segnale HR rimane invece stabile per un intervallo considerevolmente più lungo, con change point rilevati solo ai minuti 26.5 e 28. Il change point sull'EDA al minuto 17 coincide con l'inizio dell'interazione sociale, marcando con precisione il passaggio tra le due fasi della sessione. L'aspetto più rilevante ai fini della validazione è tuttavia lo sfasamento temporale tra le risposte autonome: nonostante l'intensa attività motoria iniziata al minuto 17, la frequenza cardiaca rimane stabile per quasi dieci minuti, confermando il disaccoppiamento fisiologico-comportamentale tipico delle auto-stimolazioni non stressogene. Questi dati sono stati utilizzati come proof of concept per valutare la capacità predittiva dell'algoritmo: l'attivazione simpatica precoce rilevata sull'EDA al minuto 17 anticipa la risposta cardiovascolare, registrata sull'HR al minuto 26.5, di circa dieci minuti.

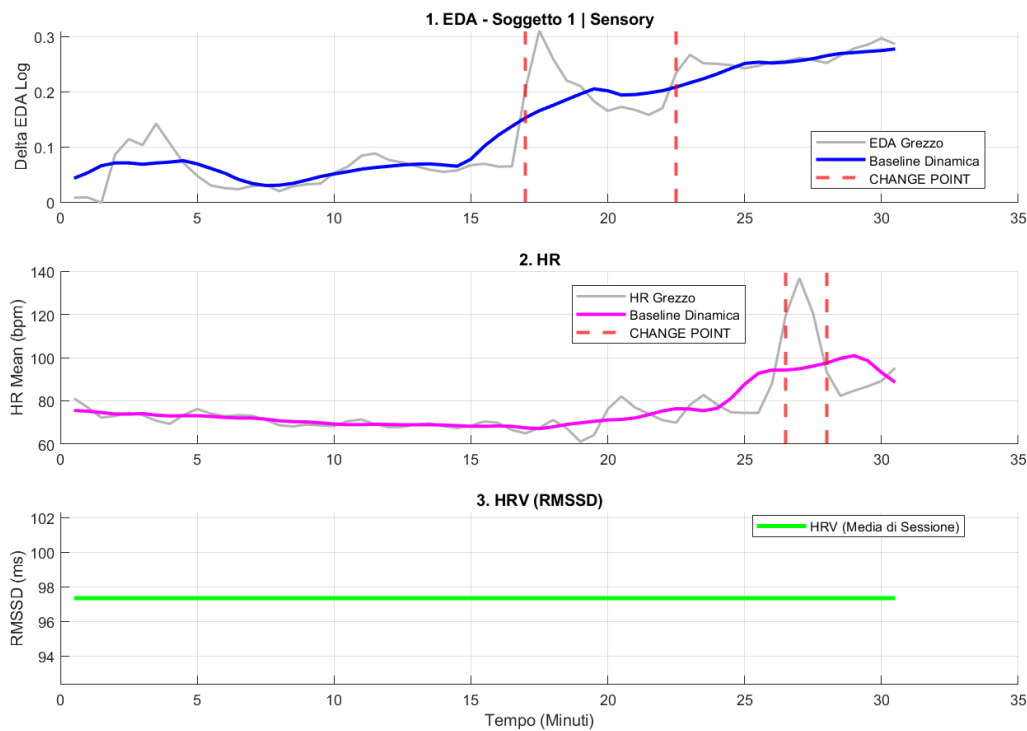


Figura 6.6: Change Point Analysis – Sessione S1_soggetto1_sensory (rilassamento iniziale seguito da interazione sociale). Dall’alto: Delta EDA log con change point ai minuti 17 e 22.5 (inizio e consolidamento dell’interazione sociale); HR mean con change point ai minuti 26.5 e 28; HRV RMSSD costante di sessione a 97 ms. Lo sfasamento di circa 9 minuti tra la prima attivazione EDA e il primo change point HR costituisce il margine di anticipo predittivo del modello.

6.1.4 Modello supervisionato: Random Forest soggetto-specifico

Il modello Random Forest descritto in questa sezione è stato sviluppato dalla collega Silvia Barbera nell’ambito del medesimo progetto di tesi. Ne viene qui presentata una sintesi funzionale perché costituisce il secondo stadio del modello ibrido descritto nel capitolo 7.

Motivazione della scelta

La Change Point Analysis descritta nella sezione 6.1 fornisce una localizzazione non supervisionata delle discontinuità strutturali nel segnale fisiologico, identificando i momenti in cui la dinamica autonoma del soggetto subisce una variazione statisticamente rilevante. Per arricchire questa informazione con una classificazione esplicita del rischio, è stato sviluppato un secondo modello di natura supervisionata, basato sul

Random Forest, che apprende dalla storia fisiologica individuale e assegna a ciascuna finestra temporale una probabilità di pre-crisi.

Architettura della pipeline

Fase A: definizione del ground truth clinico. In assenza di etichette cliniche certificate, il ground truth è derivato tramite regole soglia personalizzate per ciascun soggetto, fondate sulla letteratura scientifica di riferimento:

- **Soggetto 1:** il modello è fondato sulla teoria polivagale di Porges [9] e sul modello di rischio di Fenning et al. [5]. Le risposte di orientamento all'interazione sociale, descritte come *social context switch* [6], sono tollerate dal sistema: l'interazione tra il soggetto e l'operatore genera un'elevata attivazione simpatica regolata da un HRV elevato, configurazione che il modello riconosce come fisiologica. L'allarme si attiva solo in presenza di un crollo vagale non compensato: *Phasic_max* supera l'85° percentile e *HRV_RMSSD* scende al di sotto del 90% della mediana individuale.
- **Soggetto 2:** il modello integra il paradigma del paradosso dell'oggetto [29] e la teoria dell'interferenza [1]. I picchi di conduttanza non vengono penalizzati se accompagnati da un corretto stimming motorio rilevato sull'ACC. L'allarme si attiva esclusivamente in presenza di un disaccoppiamento autonomico-motorio: *Phasic_max* supera l'85° percentile individuale e *ACC_mean* scende al di sotto dell'80% della mediana individuale. L'interruzione di un'azione ripetitiva, come l'interazione con le luci osservata in questa sessione, può generare alterazioni emotive in un soggetto ASD che utilizza il movimento come strategia di autoregolazione.
- **Soggetto 3:** il modello è fondato sull'evidenza dell'*anticipatory rise* e del carico allostatico [30]. Il sistema valuta il rischio non solo attraverso il picco fasico, ma attraverso la pendenza tonica cumulativa del segnale, utilizzando le sessioni di lavoro cognitivo come baseline di regolazione attiva. L'allarme scatta se *Phasic_max* supera il 95° percentile calcolato sulla sessione work e *EDA_slope* supera il 75° percentile individuale, ovvero solo quando il picco è il culmine

di una rapida escalation non regolata. Questa logica permette di distinguere la concentrazione attiva dal sovraccarico cognitivo che precede l'attacco di panico.

Come salvaguardia, qualora il numero di eventi etichettati risultasse inferiore a due, le finestre corrispondenti al 10° percentile superiore di `Phasic_max` vengono automaticamente considerate eventi di rischio.

Fase B: costruzione del target predittivo e finestra di pre-allarme. Il vettore `target_warning_phase` è costruito mediante `time-shifting` a ritroso di 3 step (equivalenti a 1.5 minuti, dato che ogni finestra ha ampiezza di 30 secondi). Questo consente al modello di apprendere i precursori fisiologici dell'evento e non l'evento stesso, rendendolo utilizzabile in un contesto di allerta precoce.

Fase C: addestramento e cross-validazione. Il classificatore è addestrato tramite `fitcensemble` con metodo `Bagging` su 50 alberi decisionali (`templateTree`, `MinLeafSize = 1`), con cross-validazione `K-Fold` ($K = 5$). Le feature fisiologiche fornite in ingresso sono undici: `HR_Mean`, `HR_Std`, `Delta_HR`, `EDA_Tonic_Log`, `Delta_EDA_Log`, `EDA_Slope`, `Phasic_Max`, `Phasic_AUC`, `Driver_Peaks`, `ACC_Mean`, `HRV_RMSSD`.

Fase D: validazione visiva guidata dal diario clinico. Il modello finale, addestrato sull'intero dataset soggetto-specifico, viene applicato alla sessione clinicamente più rilevante per ciascun soggetto per produrre una curva di probabilità di rischio nel tempo. Viene calcolato il `lead time`, ovvero l'anticipo temporale con cui il modello supera la soglia di allarme (fissata al 25%) rispetto al picco fisiologico massimo della sessione.

Risultati e validazione clinica

Le metriche di valutazione estratte dalla cross-validazione `K-Fold` ($K = 5$) sono riportate in Tabella 6.1. La validazione visiva mostra la capacità del modello di generare un segnale di pre-allarme con anticipo misurabile rispetto al picco fisiologico massi-

mo della sessione, confermando la plausibilità clinica delle predizioni in relazione agli eventi documentati nel diario clinico.

Tabella 6.1: Metriche di valutazione del modello Random Forest soggetto-specifico (K-Fold, $K = 5$).

Soggetto	Accuratezza (%)	Sensibilità (%)	Specificità (%)	AUC
Soggetto 1	95.35	83.33	98.53	0.98
Soggetto 2	95.21	88.89	96.64	0.97
Soggetto 3	99.49	85.71	100.00	1.00

7. Modello ibrido per la predizione di crisi

Il presente capitolo descrive l'integrazione di due modelli sviluppati separatamente, un modello non supervisionato di anomaly detection con Change Point Analysis e un classificatore supervisionato Random Forest, in un unico sistema ibrido per la previsione delle crisi in soggetti con ASD. L'obiettivo è sfruttare i punti di forza dei due approcci: il primo rileva automaticamente variazioni anomale nel segnale EDA senza richiedere etichette a priori; il secondo classifica ogni campione sulla base di pattern appresi da un paziente noto, producendo una probabilità continua di pre-crisi.

7.1 Architettura del modello ibrido

Il sistema adotta un'architettura a due stadi sequenziali che riceve in ingresso tre segnali fisiologici continui: EDA, HRV e ACC. La scelta di questo set multimodale è supportata dagli studi di Goodwin et al. [3], che hanno dimostrato come l'analisi combinata di questi segnali, acquisiti tramite biosensori da polso, permetta di prevedere episodi di aggressività e crisi comportamentali con AUC medio di 0.84, fornendo un anticipo clinicamente utile di 1–3 minuti.

Il primo stadio è il modello non supervisionato di anomaly detection: analizza il segnale EDA per individuare automaticamente il momento in cui il suo andamento si discosta dalla baseline individuale, tramite una media mobile di pre-elaborazione e un algoritmo di Change Point Analysis. Il cambiamento rilevato viene poi sottoposto a una validazione clinica su tutti e tre i segnali, confermando l'anomalia solo quando più biomarcatori concordano. Il secondo stadio è il classificatore Random Forest, addestrato sui dati del paziente noto: riceve i segnali elaborati e produce per ogni campione una probabilità continua di pre-crisi; quando questa probabilità supera la soglia del 25% in un momento precedente all'insorgenza della crisi, il sistema lancia l'allarme e calcola il lead time disponibile per l'intervento.

L'approccio a cascata (Change Point Detection seguito da classificazione supervisionata) è conforme all'architettura proposta da Apostol et al. [31], che dimostra

come l'integrazione della Change Point Detection riduca drasticamente i falsi positivi rispetto all'uso del solo classificatore su serie temporali acquisite in continuo. Tale metodologia ibrida trova riscontro anche in Muke & Kozierekiewicz [32], che combinano metodi non supervisionati e Random Forest per la classificazione del carico cognitivo da segnali fisiologici multimodali.

7.2 Generazione dei dati sintetici

Poiché nel presente lavoro non erano disponibili dati fisiologici reali contenenti episodi di aggressività, i segnali sono stati generati tramite la funzione MATLAB `genera_sessione`, progettata per simulare il comportamento fisiologico tipico di un soggetto con ASD durante una sessione di monitoraggio. La scarsità di eventi comportamentali nei dataset reali è un limite riconosciuto da Imbiriba et al. [14], che dichiarano la necessità di tecniche di bilanciamento delle classi per compensare la rarità delle crisi registrate.

Questa scelta metodologica è coerente con un approccio consolidato in letteratura. Lange et al. [33] hanno dimostrato che addestrare classificatori su dati sintetici da sensori indossabili (EDA, HRV, ACC) migliora le performance del modello dell'11.9–15.5% rispetto alla baseline su dati reali insufficienti. In modo analogo, Ismail Fawaz et al. [34] hanno mostrato che la data augmentation su serie temporali aumenta significativamente l'accuratezza dei classificatori in scenari con dataset ridotti. Infine, Sabry et al. [35] documentano che addestrare su dati sintetici di biosegnali da sensori indossabili comporta una riduzione di accuratezza massima del 12% rispetto ai dati reali, un limite accettabile nel nostro caso, giustificato dall'assenza di alternative cliniche.

7.2.1 Struttura della sessione simulata

Ogni sessione è composta da 1000 campioni con frequenza di campionamento di un campione ogni 30 secondi, per una durata totale di 500 minuti. Le quattro variabili fisiologiche sono generate come segue:

- **EDA** (`Phasic_max`): la baseline è generata tramite una distribuzione di Bates (media di 5 variabili uniformi casuali, moltiplicata per 3 e traslata di 1) che si

avvicina a una gaussiana senza produrre valori negativi, rendendola adatta a simulare un segnale fisiologico di base. Sopra la baseline vengono aggiunte tre componenti in sequenza: 30 semi-onde sinusoidali SCR spontanei prima dell'onset, una rampa esponenziale $\exp(\text{linspace}(0,2))$ normalizzata tra 0 e 1 e scalata di 8 μS nella finestra pre-crisi, e un picco di crisi lineare da 8 a 10 μS .

- **EDA slope:** calcolata come derivata discreta dell'EDA tramite la funzione `gradient`; è automaticamente nulla in baseline e massima al picco della rampa pre-crisi.
- **HRV (RMSSD):** segnale gaussiano con media 50 ms e deviazione standard 5 ms. Nella fase pre-crisi viene moltiplicato per un vettore lineare da 1 a 0.7, replicando il decremento graduale del 30% del tono parasimpatico; durante la crisi crolla al 50% del valore istantaneo.
- **ACC (ACC_mean):** segnale gaussiano con media 1.2 g (somma della forza di gravità costante a 1.0 g e dei micro-movimenti di *stimming*) e deviazione standard 0.1 g. Durante la crisi viene moltiplicato per 0.4, simulando la brusca interruzione dell'attività motoria che caratterizza l'insorgenza della crisi comportamentale acuta in soggetti con ASD [3].

Il rumore gaussiano su HRV e ACC è necessario per conferire realismo ai dati: su segnali perfettamente puliti qualsiasi classificatore raggiungerebbe performance banalmente ottimali, rendendo il test scientificamente privo di valore.

7.3 Anomaly detection e Change Point Analysis

7.3.1 Pre-elaborazione del segnale EDA

Prima dell'analisi, sul segnale EDA grezzo viene calcolata una media mobile con finestra di 5 campioni (funzione `movmean`), che attenua le fluttuazioni rapide dei picchi fasici SCR spontanei. Questo segnale levigato viene utilizzato esclusivamente come input per l'algoritmo di change point, mentre il segnale originale viene conservato per le successive fasi di classificazione.

7.3.2 Rilevamento del change point

L'algoritmo `findchangepts` viene applicato ai primi 900 campioni del segnale pre-elaborato (90% della sessione). Il taglio al 90% è motivato dalla necessità di escludere la lunga fase di discesa post-crisi (campioni 850–1000), che costituisce il cambiamento matematicamente più evidente sull'intera sessione e distrarrebbe l'algoritmo dall'identificare la zona di pre-crisi. L'algoritmo identifica il punto che massimizza la differenza tra il valore medio del segmento precedente e quello del segmento successivo, identificando così la transizione più netta dell'andamento del segnale.

7.3.3 Validazione clinica multi-segnale

Il change point rilevato viene sottoposto a una validazione clinica basata su quattro regole combinate, derivate dall'analisi dei singoli soggetti nei capitoli precedenti e dalla letteratura di riferimento [3]:

- **EDA alta:** la componente fasica supera l'85° percentile individuale, segno di iper-arousal acuto;
- **Slope EDA alta:** la pendenza EDA supera il 75° percentile individuale, segno di salita rapida del segnale;
- **Crollo motorio:** l'ACC è scesa sotto l'80% della mediana individuale, possibile blocco motorio pre-crisi;
- **Crollo vagale:** l'HRV RMSSD è scesa sotto il 90% della mediana individuale, indice di attivazione simpatica immediatamente precedente alla crisi.

La crisi viene confermata se la regola EDA è vera e almeno una tra le restanti tre è vera (struttura logica AND+OR). Questa architettura logica riduce i falsi positivi: un singolo segnale anomalo non è sufficiente alla predizione, ne servono almeno due concordi [31]. La mediana è preferita alla media per ACC e HRV perché è meno sensibile agli artefatti, frequenti nei segnali da sensori indossabili. Se nessun change point supera le regole cliniche, viene attivato un meccanismo di *fallback* che etichetta come pre-crisi i campioni con EDA superiore al 95° percentile.

7.4 Classificatore Random Forest

7.4.1 Definizione delle etichette

Per l'addestramento supervisionato, i campioni della sessione di training vengono etichettati come "pre-crisi" nei 6 campioni immediatamente precedenti alla crisi (campioni 794–799, corrispondenti ai 3 minuti di anticipo indicati da Imbiriba et al. [14] come soglia clinicamente utile per l'intervento preventivo) e come "calma" in tutti gli altri.

7.4.2 Addestramento

Il modello è un insieme di 50 alberi decisionali addestrato con il metodo Bagging (funzione MATLAB `fitcensemble`, opzione `bag`). Ogni albero viene addestrato su un campione casuale con ripetizione dei dati originali, con dimensione minima delle foglie pari a 1. Le quattro feature di ingresso sono: `Phasic_max`, `EDA_slope`, `HRV_RMSSD` e `ACC_mean`. La previsione finale è la media delle probabilità dei 50 alberi, significativamente più stabile di un singolo albero.

7.4.3 Predizione e soglia di allarme

Il modello viene applicato alla sessione del paziente ignoto (`T_test`): per ogni campione viene prodotta una probabilità di pre-crisi tra 0% e 100%. La probabilità viene levigata tramite un filtro gaussiano con finestra di 10 campioni (funzione `smoothdata`) per attenuare i picchi isolati da rumore senza perdere i picchi significativi. L'allarme viene lanciato nel momento in cui la probabilità supera la soglia del 25%, purché ciò avvenga prima dell'inizio della crisi.

7.5 Risultati del modello ibrido

7.5.1 Dati artificiali, change point e predizione

La figura 7.1 mostra i tre pannelli della sessione sintetica. Il segnale EDA presenta chiaramente le tre fasi: una baseline stabile compresa tra 2 e 5 μS con piccoli picchi SCR spontanei, una rampa di crescita esponenziale nella fase pre-crisi, e un picco massimo di circa 13 μS al momento della crisi. Il change point (linea rossa tratteggiata) viene rilevato correttamente in corrispondenza dell'inizio della fase pre-crisi, intorno al campione 400. ACC e HRV normalizzati mostrano un calo sincrono, scendendo al 50% del proprio valore basale in coincidenza con l'evento di crisi, confermando la coerenza tra i tre segnali. Il classificatore Random Forest lancia l'allarme con un anticipo di 2.5 minuti rispetto all'inizio della crisi, in linea con la finestra temporale di circa 3 minuti indicata da Imbiriba et al. [14] come soglia clinicamente utile.

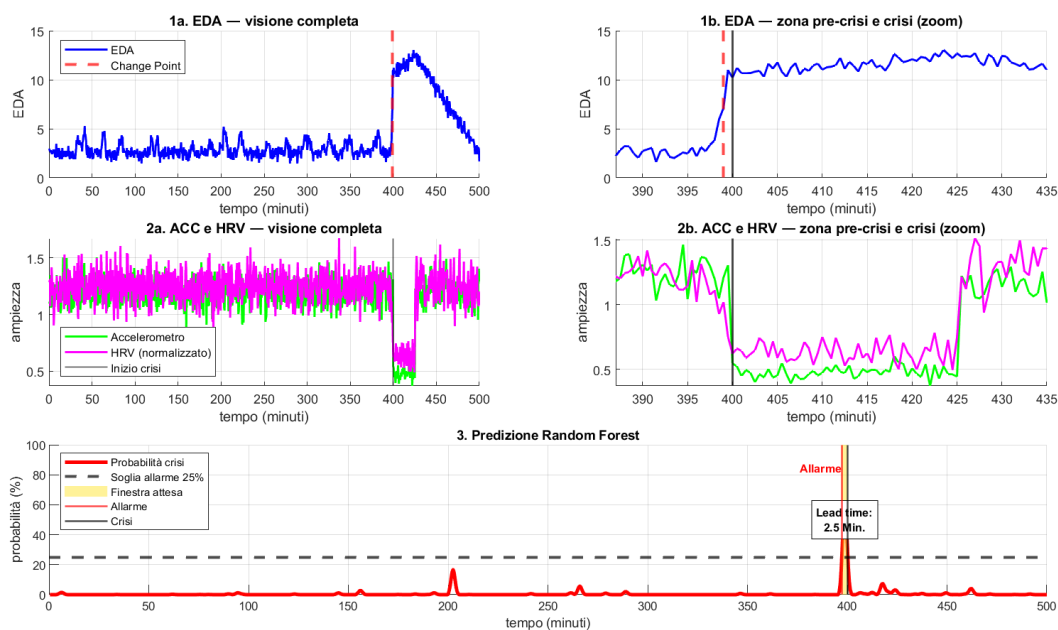


Figura 7.1: Sessione sintetica completa: (1) segnale EDA con change point rilevato; (2) ACC e HRV normalizzati con indicazione dell'inizio crisi; (3) probabilità di pre-crisi predetta dal Random Forest con soglia di allarme al 25% e lead time di 2.5 minuti.

7.5.2 Matrice di confusione

La figura 7.2 mostra la matrice di confusione sul set di test. Il classificatore ha analizzato in totale 1000 campioni: 993 campioni di calma sono stati classificati correttamente come negativi (TN), mentre 6 campioni pre-crisi sono stati rilevati correttamente (TP). Un solo campione pre-crisi non è stato rilevato (FN), corrispondente a un istante di transizione tra baseline e fase pre-crisi in cui il segnale non era ancora sufficientemente distinto dalla calma. Non si registra alcun falso positivo (FP=0). Il sistema raggiunge una sensibilità dell'85.7%, una specificità del 100% e una precisione del 100%. L'assenza totale di falsi positivi è particolarmente rilevante in ambito clinico, poiché garantisce che il caregiver non venga disturbato da allarmi inutili.

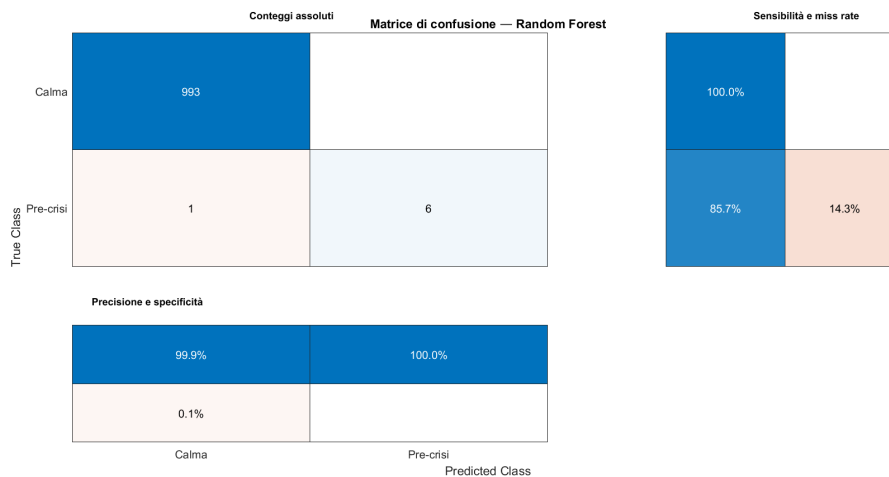


Figura 7.2: Matrice di confusione del modello ibrido sul set di test: conteggi assoluti (pannello sinistro), sensibilità e miss rate (pannello destro), precisione e specificità (pannello inferiore).

7.5.3 Curva ROC e AUC

La figura 7.3 mostra la curva ROC con $AUC=0.85$. Questo valore indica che il modello ha una probabilità dell'85% di assegnare una probabilità più alta a un campione pre-crisi rispetto a un campione di calma scelto casualmente. Valori di $AUC > 0.80$ sono considerati buoni per classificatori biomedici in letteratura [3].

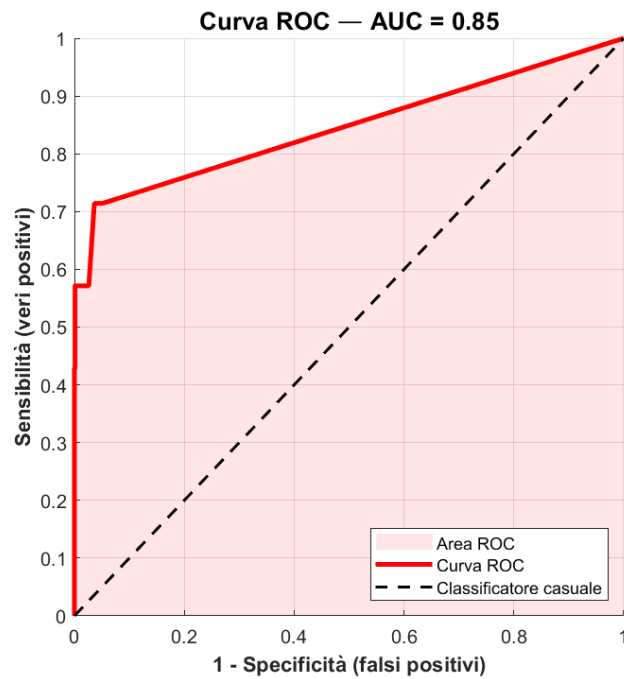


Figura 7.3: Curva ROC del classificatore Random Forest: AUC = 0.85.

7.5.4 Importanza delle feature

La figura 7.4 mostra l'importanza relativa delle quattro feature calcolata dal modello Random Forest. La feature più importante è la pendenza dell'EDA (EDA slope), con un peso circa 10 volte superiore rispetto alle altre: il modello ha appreso che il segnale più utile per distinguere la fase pre-crisi dalla calma è la velocità con cui l'EDA cresce, non il suo valore assoluto. L'HRV si posiziona al secondo posto, confermando il suo ruolo di biomarcatore del sistema nervoso autonomo in fase pre-crisi [3]. ACC e valore assoluto dell'EDA contribuiscono in misura minore e comparabile tra loro.

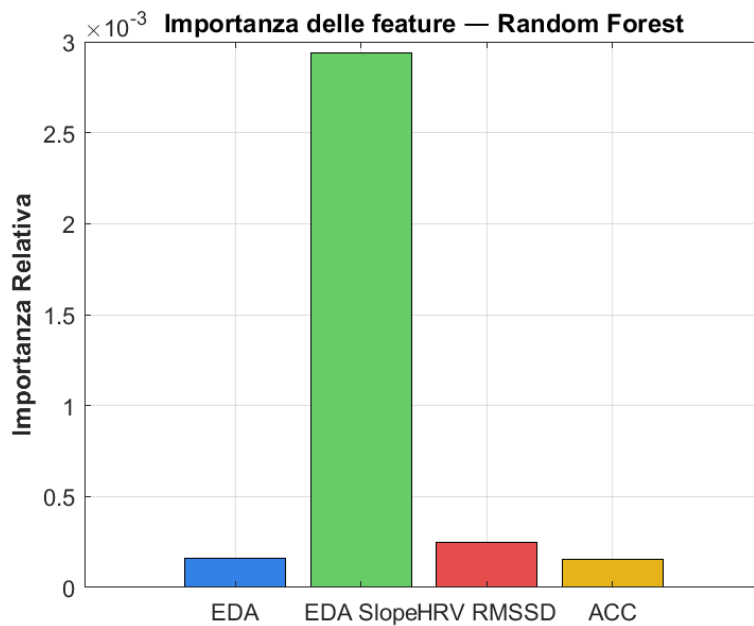


Figura 7.4: Importanza relativa delle feature calcolata dal modello Random Forest. L'EDA Slope domina nettamente con un peso circa 10 volte superiore, seguita da HRV, RMSSD, EDA e ACC.

7.5.5 Confronto training vs test

La figura 7.5 confronta la probabilità di pre-crisi predetta sul training set (paziente noto) e sul test set (paziente ignoto). Sul training set la probabilità raggiunge un picco del 68%; sul test del 58%: una differenza di soli 10 punti percentuali che indica una buona capacità di generalizzazione del modello su un paziente ignoto. Entrambi i picchi superano ampiamente la soglia del 25%. Prima della crisi, il test mostra un picco sporadico di circa il 17% intorno al minuto 200, un falso positivo che non supera la soglia e non genera allarme.

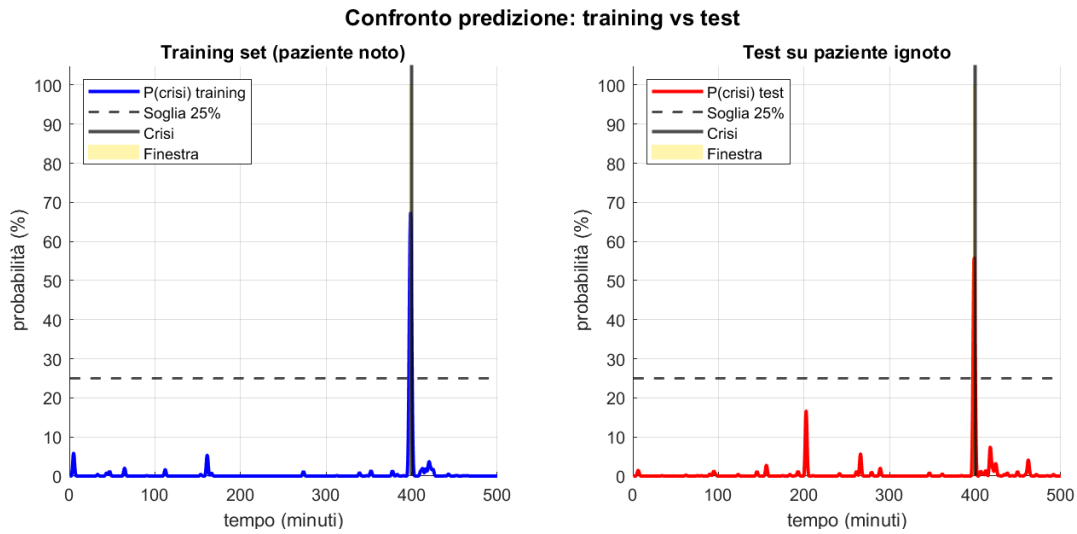


Figura 7.5: Confronto della probabilità di pre-criisi predetta dal Random Forest sul training set (blu, paziente noto) e sul test set (rosso, paziente ignoto). La soglia di allarme al 25% è indicata dalla linea tratteggiata; la finestra di pre-criisi attesa è evidenziata in giallo, anche se poco visibile.

7.6 Validazione della robustezza temporale

Per valutare la solidità del modello ibrido al variare della posizione temporale della crisi, sono state condotte due analisi di robustezza distinte: una focalizzata sullo stadio di Change Point Analysis (`hybrid_cpa.m`) e una focalizzata sullo stadio Random Forest (`hybrid_RF.m`). In entrambi i casi, la crisi è stata postata in cinque posizioni temporali diverse rispetto all'onset originale, con 10 ripetizioni indipendenti per ciascuna posizione (seed del generatore casuale variato ad ogni ripetizione).

7.6.1 Robustezza della Change Point Analysis

L'analisi di robustezza della CPA è stata condotta spostando l'onset della crisi ai campioni 680, 720, 748, 755 e 770, corrispondenti rispettivamente ai minuti 340, 360, 374, 377.5 e 385, in aggiunta allo scenario originale al minuto 400. In questo caso il parametro aggiuntivo monitorato è la percentuale di ripetizioni in cui il change point supera la validazione clinica multi-segnale, ovvero la proporzione di sessioni in cui l'anomalia EDA è confermata da almeno un altro biomarcatore nel punto di cambiamento.

Figura 6: Robustezza temporale del modello ibrido (CPA + RF)

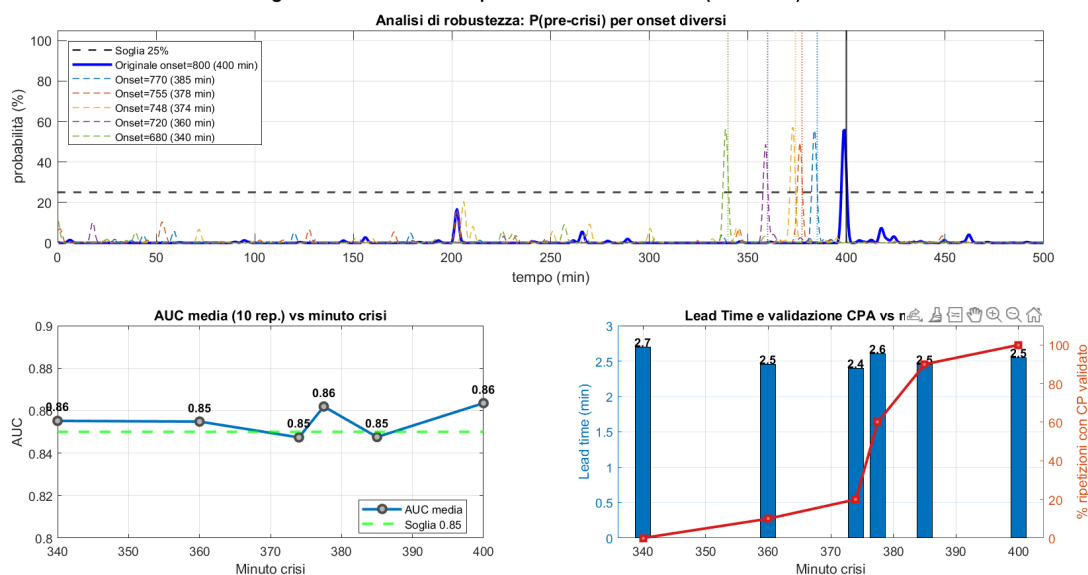


Figura 7.6: Analisi di robustezza del modello ibrido al variare dell'onset della crisi (minuti 340–400, 10 ripetizioni per scenario). Pannello superiore: $P(\text{pre-crisi})$ per onsets diversi. Pannello inferiore sinistro: AUC media e deviazione standard. Pannello inferiore destro: lead time medio (asse sinistro) e percentuale di ripetizioni con change point clinicamente validato (asse destro, curva rossa).

La figura 7.6 mostra i risultati. Il pannello superiore confronta la curva $P(\text{pre-crisi})$ originale con quelle ottenute negli scenari alternativi: ciascuna curva mostra un picco di probabilità in corrispondenza della propria crisi, dimostrando che il sistema si adatta autonomamente alla posizione temporale dell'evento. Il pannello inferiore sinistro mostra l'AUC media al variare del minuto di crisi: i valori oscillano tra 0.85 e 0.86, mantenendosi costantemente al di sopra della soglia 0.85, con deviazioni standard ± 0.04 – 0.08 , indicando stabilità robusta. Il pannello inferiore destro presenta, su doppio asse, il lead time medio (barre blu) e la percentuale di ripetizioni con CP clinicamente validato (curva rossa). Quest'ultima cresce progressivamente all'avvicinarsi dell'onset all'originale (da 0% al minuto 340 fino a 100% al minuto 385–400), indicando che la CPA valida correttamente il change point con maggiore affidabilità quando la finestra pre-crisi è temporalmente più distante dall'inizio della sessione, ovvero quando il segnale EDA ha avuto più tempo per mostrare la propria rampa esponenziale.

7.6.2 Robustezza del Random Forest

L'analisi di robustezza del Random Forest è stata condotta spostando l'onset della crisi ai campioni 300, 450, 600, 700 e 750, corrispondenti rispettivamente ai minuti 150, 225, 300, 350 e 375, in aggiunta allo scenario originale al minuto 400. Questa finestra di test più ampia rispetto all'analisi CPA è motivata dall'obiettivo di verificare la capacità del Random Forest di generalizzare su un arco temporale esteso, coprendo posizioni della crisi molto diverse dalla sessione di training.

La figura 7.7 mostra i risultati. Il pannello superiore evidenzia come il picco di probabilità segua sempre correttamente la posizione della crisi, indipendentemente dal minuto in cui essa si manifesta: ogni curva tratteggiata mostra il proprio picco allineato alla rispettiva linea verticale di crisi, a conferma che il modello risponde al segnale fisiologico e non a pattern temporali fissi. Il pannello inferiore sinistro mostra l'AUC media per i sei scenari: i valori oscillano tra 0.84 e 0.86, con una lieve flessione a 0.84 per gli onset più distanti (225 min e 350 min), rimanendo tuttavia al di sopra del riferimento 0.84 riportato da Goodwin et al. [3]. Il pannello inferiore destro mostra il lead time medio per ciascun scenario: i valori si mantengono nell'intervallo 2.4–2.6 minuti per tutti gli onset testati, confermando che l'anticipo fornito dal sistema è stabile e clinicamente utile indipendentemente dal momento della giornata in cui la crisi si manifesta [14].

Figura 6: Analisi di robustezza temporale

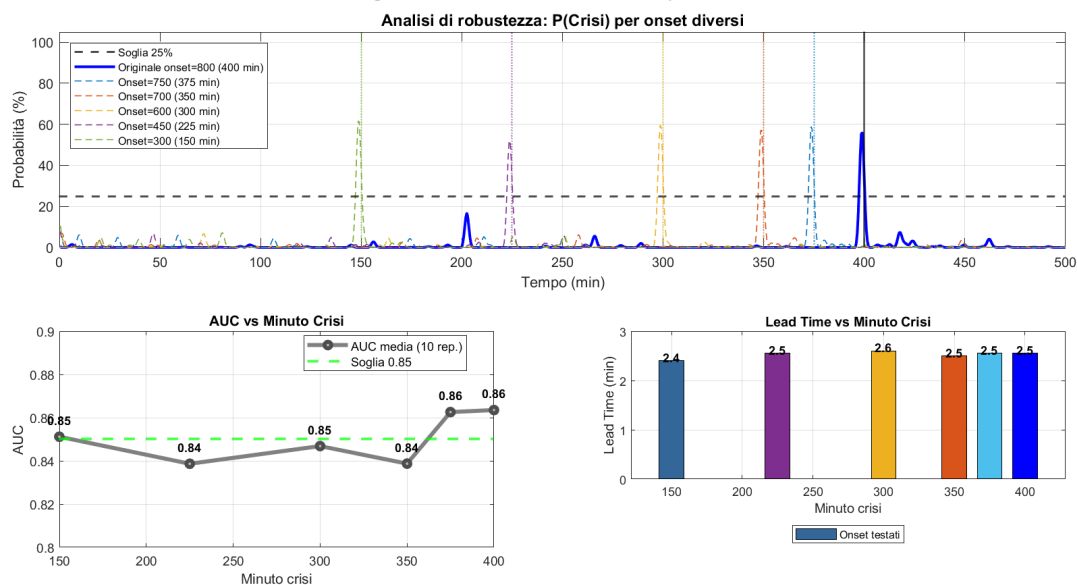


Figura 7.7: Analisi di robustezza del Random Forest al variare dell'onset della crisi (minuti 150–400, 10 ripetizioni per scenario). Pannello superiore: $P(\text{Crisi})$ per onset diversi. Pannello inferiore sinistro: AUC media per ciascun minuto di crisi. Pannello inferiore destro: lead time medio per ciascun scenario.

7.7 Analisi di robustezza al rumore

7.7.1 Il rumore nei segnali fisiologici

Nei dispositivi indossabili, il segnale registrato non è mai perfettamente pulito. Le interferenze che degradano la qualità acquisita derivano principalmente da due fattori: gli artefatti da movimento, causati dallo sfregamento del braccialetto sulla pelle, da perdite temporanee di contatto del sensore o da movimenti bruschi del braccio, e le microfluttuazioni fisiologiche, ovvero piccole variazioni basali del corpo legate ad esempio alla termoregolazione, prive di significato clinico rispetto allo stress o alle crisi. Nel modello sviluppato, queste interferenze sono state simulate aggiungendo componenti stocastiche al segnale pulito, così da ottenere un paziente virtuale fisiologicamente realistico.

7.7.2 Il test di stress a +6 dB

Nel modello di partenza, il rumore era stato calibrato su valori moderati, adatti a dimostrare il corretto funzionamento del sistema ibrido in condizioni ideali. In particolare, l'EDA era generata con ampiezza massima di 3 μ S, l'HRV con deviazione standard di 5 ms e l'ACC con deviazione standard di 0.1 g. Per valutare la resistenza dell'architettura al degrado del rapporto segnale-rumore (SNR) tipico delle condizioni d'uso reale, è stato condotto un test di stress raddoppiando l'ampiezza del rumore di ciascun segnale, il che corrisponde a un incremento di +6 dB. La relazione tra ampiezza e potenza del rumore implica che, raddoppiando l'ampiezza, la potenza quadruplica e l'SNR si riduce esattamente di 6 dB secondo la definizione:

$$\text{SNR} = 20 \cdot \log_{10} \left(\frac{\mu}{\sigma} \right)$$

dove μ è il valore medio del segnale e σ è la deviazione standard del rumore, usati come stima rispettivamente dell'ampiezza del segnale utile e dell'ampiezza del disturbo. Questa formulazione è equivalente alla definizione in potenza $10 \cdot \log_{10}(P_s/P_n)$ poiché la potenza è proporzionale al quadrato dell'ampiezza. Nel modello ideale l'SNR si attestava intorno a +20 dB (20.0 dB per HRV, 21.6 dB per ACC); con il raddoppio del rumore scende di 6 dB in entrambi i segnali. I parametri aggiornati per il test di stress sono diventati: EDA con ampiezza massima 6 μ S, HRV con deviazione standard 10 ms e ACC con deviazione standard 0.2 g.

7.7.3 Risultati del test di stress

Segnali e change point

I segnali con rumore raddoppiato appaiono visibilmente caotici rispetto al modello di partenza (figura 7.8): l'EDA presenta picchi casuali che già in fase basale sfiorano gli 8 μ S, mentre HRV e ACC mostrano oscillazioni molto più marcate, simulando un paziente con movimenti frequenti e un sensore soggetto a sfregamento continuo. Nonostante il degrado, la media mobile a 5 campioni applicata nello stadio 2 del modello ha attenuato le fluttuazioni ad alta frequenza, permettendo all'algoritmo `findchangepts`

di individuare la transizione strutturale esattamente al minuto 400 senza alcun ritardo: il primo stadio del modello ibrido si è dimostrato robusto al livello di degrado testato.

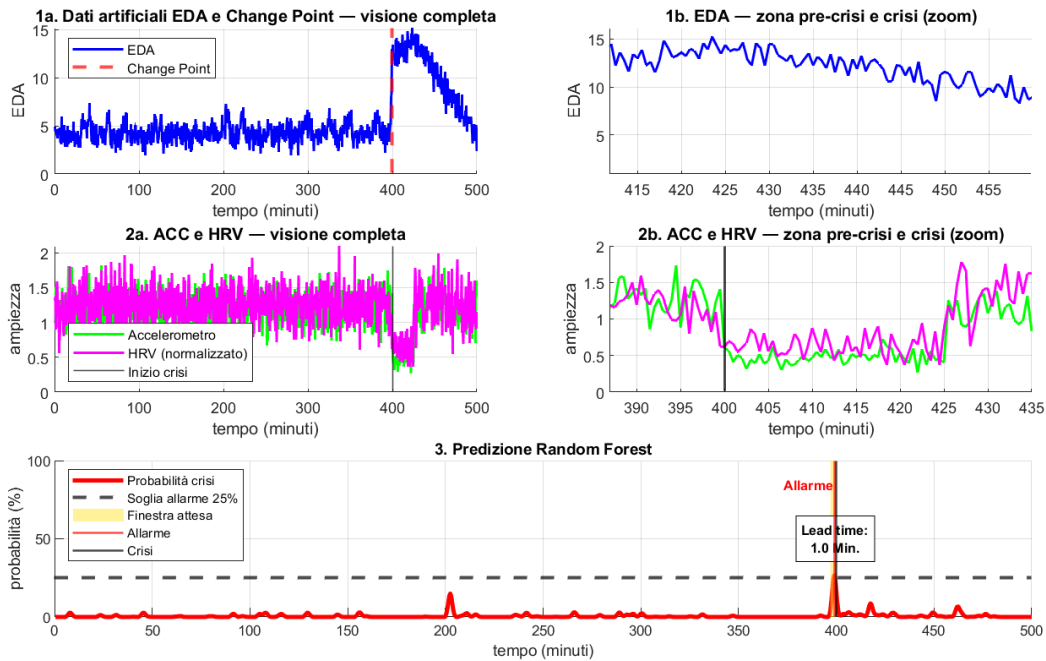


Figura 7.8: Test di stress +6 dB: (1) segnale EDA con change point rilevato; (2) ACC e HRV normalizzati; (3) probabilità di pre-crisi con allarme al minuto 399 e lead time di 1.0 minuto.

Crollo del lead time e della sensibilità

Il comportamento del Random Forest ha invece subito un degrado significativo. Nel modello originale la curva di probabilità rimaneva stabile a zero per quasi tutta la sessione, con un unico piccolo picco intorno al minuto 210, e lanciava l'allarme con 2.5 minuti di anticipo: il segnale pulito consentiva al classificatore di rilevare i microcambiamenti fin dalle prime fasi della rampa di accumulo. Con il rumore a +6 dB la curva non è più piatta: presenta fluttuazioni continue e intorno al minuto 200 genera un picco che sfiora il 20%; il lead time si è ridotto a 1.0 minuto, con una perdita netta di 1.5 minuti di finestra predittiva.

La figura 7.9 mostra l'effetto sulla matrice di confusione. Nel modello originale i veri positivi erano 6 su 7 (sensibilità 85.7%); nel test di stress a +6 dB i veri positivi si riducono a 1 su 7, con la sensibilità crollata al 14.3%: il classificatore ha scambiato per calma quasi tutta la fase di pre-crisi, riconoscendo l'anomalia solo nell'ultimo

campione, circa 0.5 minuti prima dell'evento. La specificità si è mantenuta al 100% in entrambi i casi, senza falsi allarmi durante le fasi di calma.

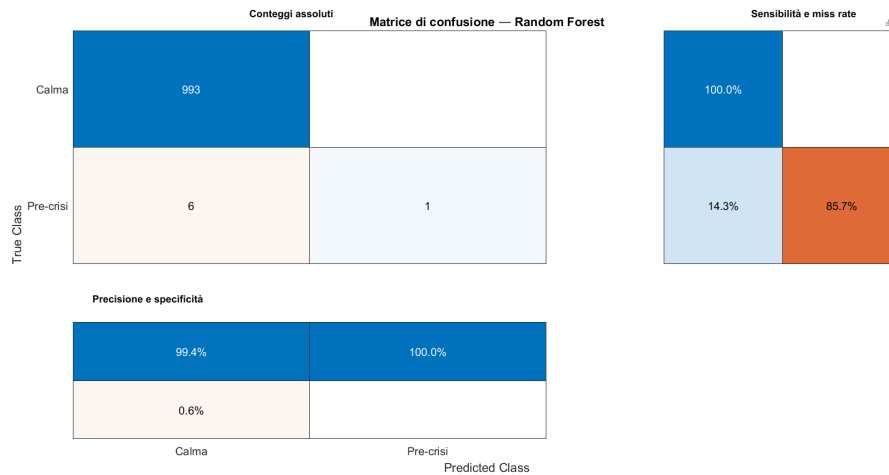


Figura 7.9: Confronto della matrice di confusione tra modello originale (sinistra) e test di stress +6 dB (destra): la sensibilità crolla dall'85.7% al 14.3%, mentre la specificità rimane al 100% in entrambi i casi.

Curva ROC e importanza delle feature

La curva ROC (figura 7.10) riflette il degrado: l'AUC scende da 0.85 a 0.76. Il modello mantiene una discreta capacità discriminativa globale, ma per mantenere lo stesso anticipo del modello originale sarebbe necessario abbassare la soglia di allarme con il rischio di un numero inaccettabile di falsi positivi.

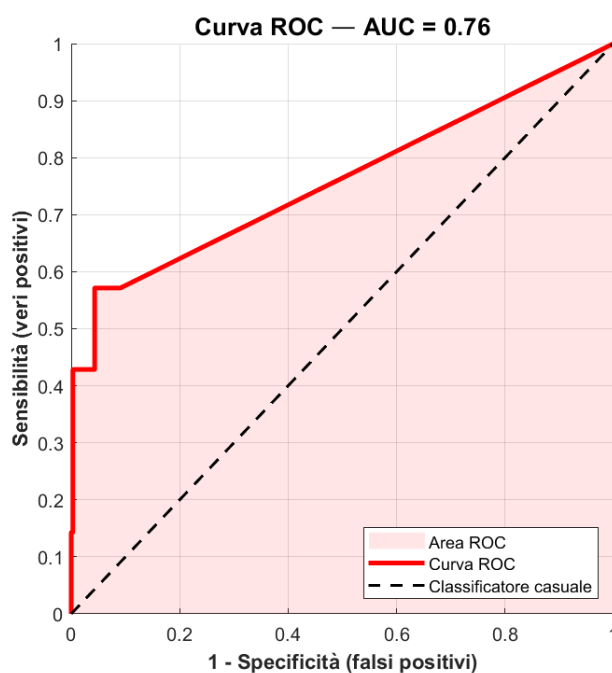


Figura 7.10: Curva ROC con +6 dB di rumore: AUC = 0.76, in calo rispetto al valore originale di 0.85.

L'analisi dell'importanza delle feature (figura 7.11) rivela la causa del degrado. Nel modello originale la pendenza dell'EDA dominava con un valore di circa 3×10^{-3} , quasi tre volte superiore a tutte le altre variabili: il classificatore aveva appreso che il predittore più utile non è la quantità di sudore in un dato istante, ma la velocità con cui sta aumentando, ovvero la rampa di accumulo che garantisce i 2.5 minuti di anticipo. Con il rumore a +6 dB, l'importanza dell'EDA slope si riduce di oltre il 60% (a circa 1.1×10^{-3}) e i pesi si ridistribuiscono in modo più uniforme sulle altre variabili. La causa è matematica: il calcolo della pendenza equivale a derivare numericamente il segnale, e la derivazione amplifica le fluttuazioni ad alta frequenza introdotte dal rumore, rendendo la variabile caotica e non informativa. Il modello si basa quindi sui dati grezzi di EDA, ACC e HRV, che però reagiscono solo a ridosso dell'evento, con conseguente perdita della finestra di pre-allarme.

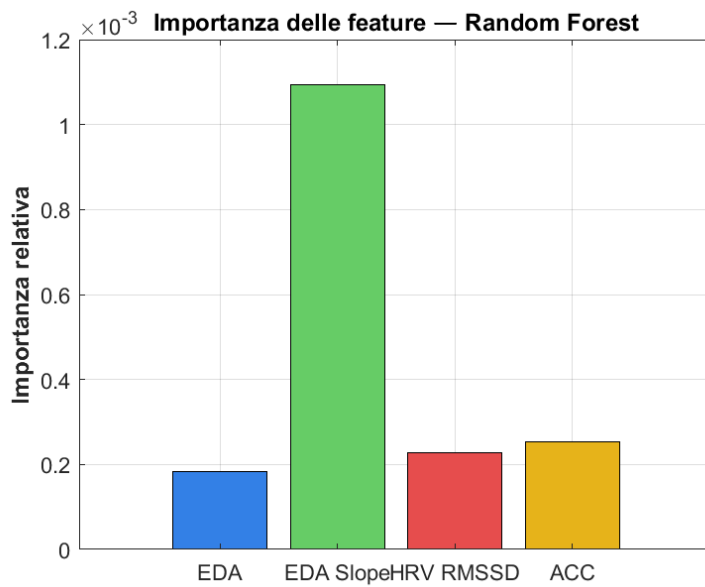


Figura 7.11: Importanza delle feature con rumore +6 dB: il peso dell'EDA slope si riduce di oltre il 60% rispetto al modello originale, con redistribuzione uniforme su EDA, ACC e HRV.

Generalizzazione con rumore elevato

La figura 7.12 rivela la perdita di capacità di generalizzazione in condizioni di rumore elevato. Sui dati di addestramento il classificatore mantiene un picco di circa il 57% alla crisi, riconoscendo correttamente l'evento anche con il segnale degradato. Sui dati di verifica la curva di probabilità supera a malapena la soglia operativa del 25% nell'istante di crisi, con un picco spurio di circa il 15% intorno al minuto 200 durante la baseline: le fluttuazioni continue causate dagli artefatti simulati riducono l'affidabilità del classificatore sui campioni mai visti. Questo risultato motiva direttamente gli interventi di pre-processing descritti nella sezione successiva.

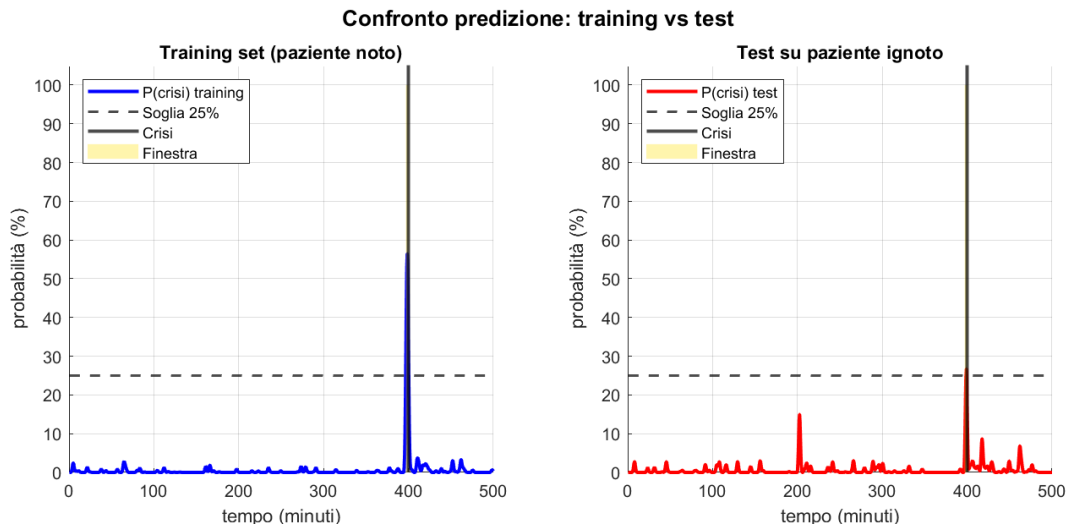


Figura 7.12: Confronto training vs test con rumore +6 dB: il training mantiene un picco del $\sim 57\%$ alla crisi; il test supera appena la soglia del 25%, con un picco del $\sim 15\%$ al minuto 200.

7.8 Soluzione al degrado di sensibilità

7.8.1 Interventi implementati

L'analisi del crollo di sensibilità da 85.7% a 14.3% ha evidenziato che la causa principale non risiede nell'architettura del classificatore, ma nella qualità della feature più importante: l'EDA Slope. Sono state apportate due modifiche alla funzione di generazione dei dati nel codice `hybrid_noise_robust.m`.

Modifica 1: pre-processing della slope (filtro media mobile). Con rumore a +6 dB, il segnale EDA oscilla rapidamente da un campione all'altro. Quando si calcola la derivata numerica tramite `gradient` su un segnale del genere, non si ottiene la pendenza del trend generale ma la differenza tra un'oscillazione casuale e quella successiva, anch'essa casuale: la slope risultante è rumore puro. La soluzione consiste nel levigare l'EDA prima di derivarla tramite una media mobile a 5 campioni:

```
EDA_smooth = movmean(EDA, 5);
EDA_Slope = gradient(EDA_smooth);
```

Una finestra di 5 campioni, corrispondente a 2.5 minuti con la frequenza di campionamento adottata, attenua i picchi isolati di rumore, lasciando visibile il solo trend lento (ovvero la rampa esponenziale pre-crisi) perché la finestra è più corta della rampa stes-

sa. Questa finestra è già utilizzata nel Change Point Detection ($\text{movmean}(\text{Phasic_max}, 5)$), garantendo coerenza tra i due stadi del modello.

Modifica 2 – Ampiezza della rampa pre-crisi variabile. L'ampiezza della rampa esponenziale è stata resa casuale nell'intervallo 4–10 μS invece di essere fissa a 8 μS :

$$\text{amp_rampa} = 4 + 6 * \text{rand}();$$

L'intensità dell'accumulo di stress fisiologico varia tra episodi diversi e tra pazienti diversi. Questa modifica rende il problema di classificazione più realistico, evitando che il modello memorizzi un'ampiezza fissa e migliorando la capacità di generalizzazione del Random Forest su sessioni mai viste.

7.8.2 Risultati del modello robusto

Con le due modifiche introdotte, il lead time torna a 2.5 minuti, identico al modello in condizioni ideali. La figura 7.13 mostra la coerenza del risultato: la probabilità di crisi rimane al di sotto del 5% per l'intera baseline, con picchi isolati mai superiori al 15%, supera la soglia del 25% al minuto 397.5 e raggiunge il picco in corrispondenza esatta della crisi al minuto 400.

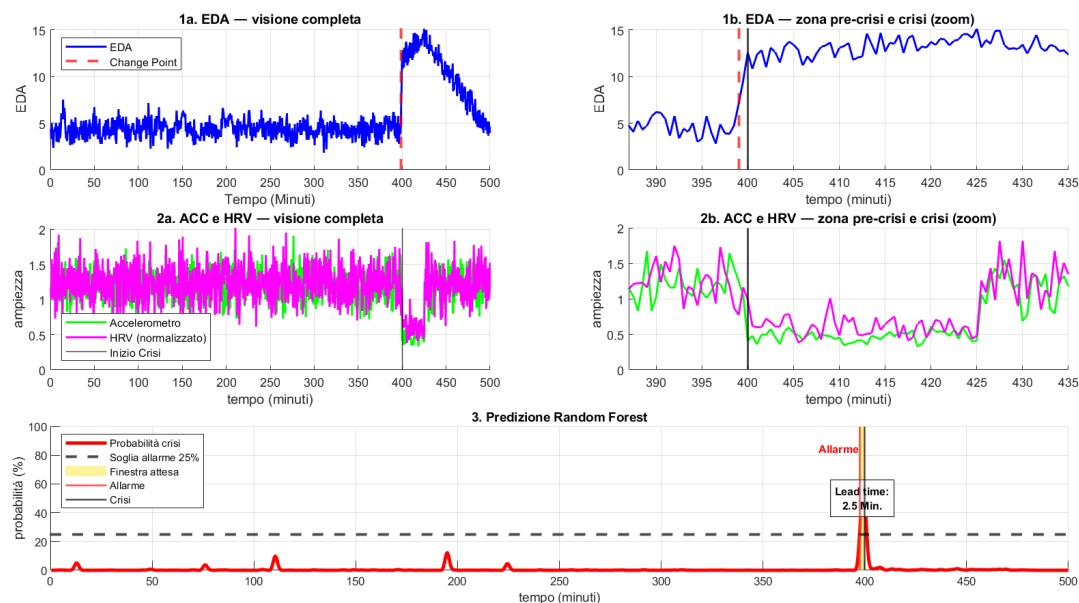


Figura 7.13: Modello robusto al rumore +6 dB: la probabilità di pre-crisi supera la soglia del 25% al minuto 397.5, con lead time di 2.5 minuti identico al modello ideale.

La figura 7.14 mostra il recupero della sensibilità: i veri positivi tornano a 6 su 7 (sensibilità 85.7%), con un solo falso positivo su 993 campioni di calma (specificità 99.9%). La lieve riduzione di specificità rispetto al modello originale (100%) è attribuibile all'introduzione della variabilità di ampiezza della rampa, che rende il confine di classificazione leggermente più conservativo.

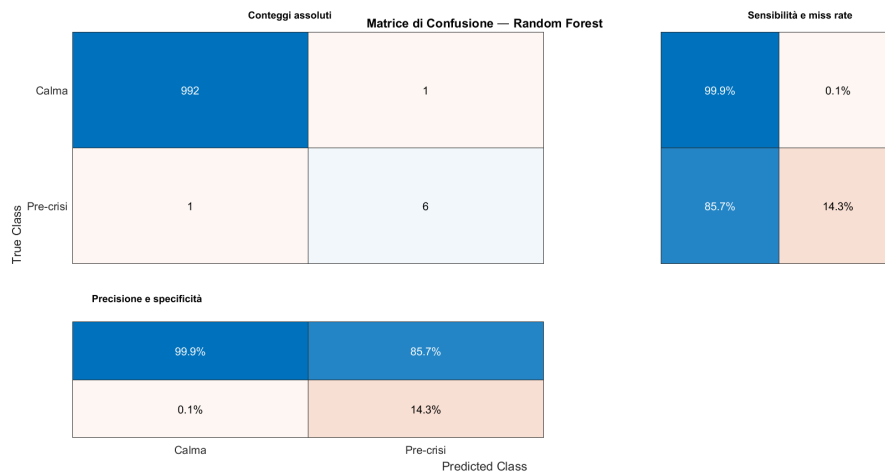


Figura 7.14: Matrice di confusione del modello robusto: sensibilità 85.7% e specificità 99.9%.

La curva ROC (figura 7.15) raggiunge $AUC = 0.91$, con una forma a gradini tipica dei dataset con pochi esempi positivi (7 campioni pre-crisi su 1000): il gradino iniziale a circa 0.86 di sensibilità con quasi zero falsi positivi indica che 6 dei 7 campioni pre-crisi vengono classificati correttamente con alta confidenza già prima di qualsiasi compromesso sulla specificità. Il valore di $AUC = 0.91$ è dato dal fatto che sono stati usati dati sintetici; su dati reali i valori si allineerebbero ai riferimenti di Goodwin et al. [3] ($AUC \approx 0.83-0.87$).

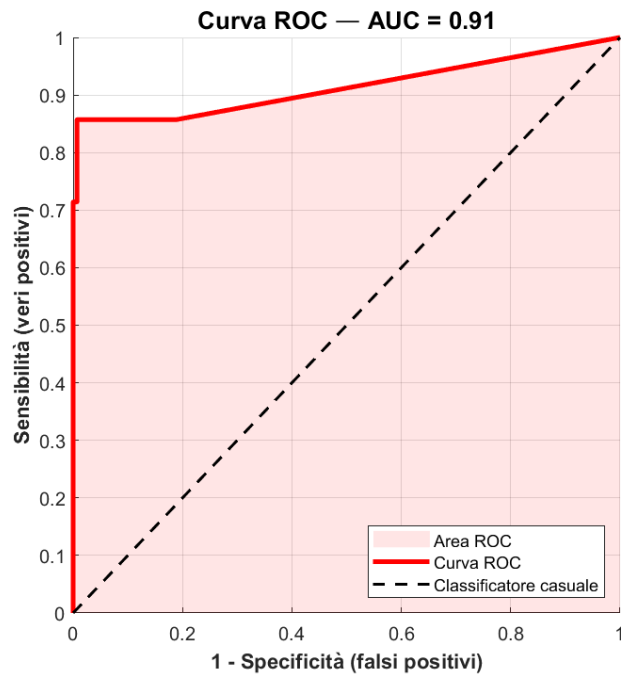


Figura 7.15: Curva ROC del modello robusto: AUC = 0.91.

La figura 7.16 mostra il ripristino dell'importanza dell'EDA slope come feature dominante. Grazie al pre-processing con media mobile, il suo peso raggiunge $\sim 5.2 \times 10^{-3}$, circa 50 volte superiore a tutte le altre variabili ($\sim 0.1 \times 10^{-3}$): un risultato superiore al modello originale in condizioni ideali ($\sim 3 \times 10^{-3}$, ~ 10 volte di prevalenza). La levigatura del segnale prima della derivazione rende il predittore non solo più stabile e interpretabile, ma anche più informativo rispetto alla versione senza pre-processing, perché elimina le micro-oscillazioni presenti nel segnale originale.

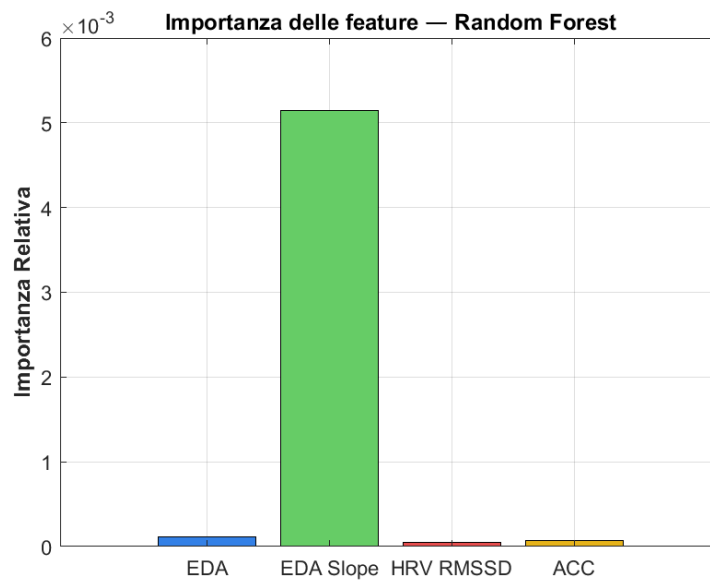


Figura 7.16: Importanza delle feature del modello robusto: l'EDA slope raggiunge $\sim 5.2 \times 10^{-3}$, circa 50 volte superiore alle altre variabili ($\sim 0.1 \times 10^{-3}$), grazie al pre-processing con media mobile prima della derivazione.

La figura 7.17 conferma il recupero della capacità di generalizzazione. Entrambe le curve superano nettamente la soglia del 25% al momento della crisi: il training raggiunge un picco di circa il 60%, il test di circa il 53%, con una differenza di soli 7 punti percentuali. Durante la baseline, entrambe le curve mostrano piccole fluttuazioni mai superiori al 12%, che non generano falsi allarmi clinicamente rilevanti. Il contrasto con il modello noisy non corretto, dove il test faticava a raggiungere il 25%, dimostra che il pre-processing della slope è sufficiente a ripristinare la generalizzazione su pazienti mai visti durante l'addestramento.

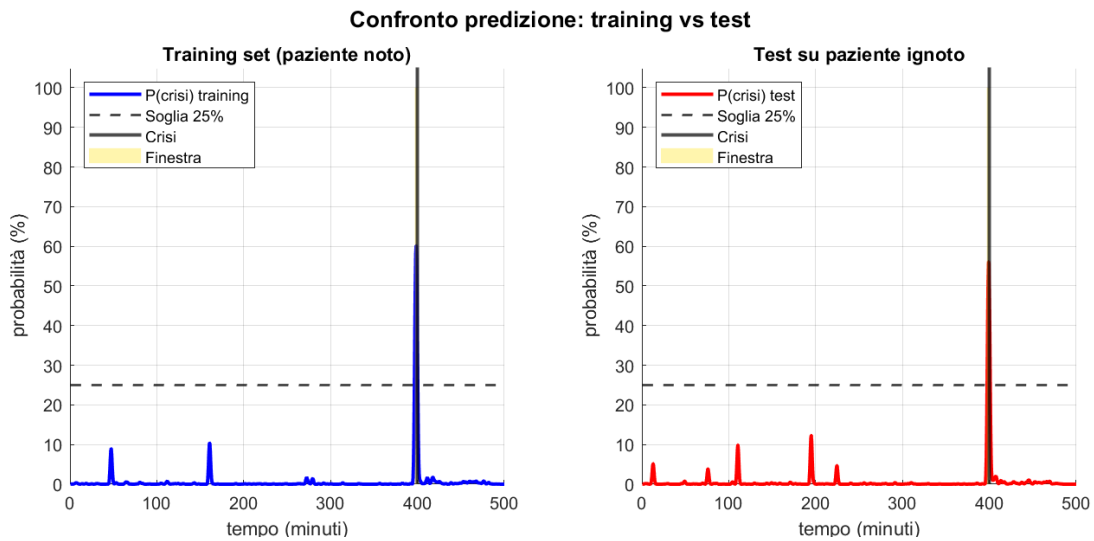


Figura 7.17: Confronto training vs test del modello robusto: training $\sim 60\%$ e test $\sim 53\%$ al momento della crisi.

La tabella 7.1 riassume le metriche di prestazione nei tre scenari analizzati.

Tabella 7.1: Confronto delle metriche di prestazione: modello ideale, test di stress +6 dB e modello robusto +6 dB.

Metrica	Modello ideale	Noisy +6 dB	Robusto +6 dB
Sensibilità	85.7%	14.3%	85.7%
Specificità	100%	100%	99.9%
Falsi allarmi	0	0	1/993
AUC	0.85	0.76	0.91
Lead Time	2.5 min	1.0 min	2.5 min

L'obiettivo di questa analisi non è massimizzare il valore assoluto dell'AUC, ma dimostrare che un intervento mirato sulla qualità delle feature, senza modificare l'architettura del classificatore, è sufficiente a ripristinare la capacità predittiva degradata dal rumore.

7.9 Prospettive di trasferibilità clinica: dai dati sintetici al contesto reale

Sebbene il modello ibrido abbia dimostrato ottime capacità predittive sui dati sintetici generati appositamente, il passaggio all'utilizzo clinico su pazienti reali tramite dispositivi indossabili, come il braccialetto Empatica E4, rappresenta una sfida complessa. Studi precedenti hanno dimostrato che segnali di EDA, HRV e ACC acquisiti da biosensori da polso permettono di predire episodi di aggressività in soggetti ASD con AUC fino a 0.84 [3], risultato esteso su larga scala a 70 pazienti in quattro ospedali con un anticipo predittivo di 3 minuti [14]: questi lavori costituiscono il riferimento metodologico e lo standard di confronto per le performance del nostro sistema. Tuttavia, lo scostamento tra l'ambiente di simulazione e il contesto reale rende necessari adeguamenti strutturali e algoritmici che vengono descritti nelle sottosezioni seguenti [32]. L'architettura logica di base, ovvero la Change Point Analysis seguita dal Random Forest, rimarrà invariata.

7.9.1 Pre-elaborazione avanzata: scomposizione dell'EDA e filtraggio degli artefatti

Nel codice attuale, l'attività elettrodermica viene elaborata calcolandone la media mobile e la derivata. Nel contesto reale, tuttavia, l'EDA è la sovrapposizione di due fenomeni distinti: il livello tonico (*Skin Conductance Level, SCL*), che varia lentamente in risposta a stimoli termici o metabolici, e le risposte fasiche (*Skin Conductance Response, SCR*), i picchi rapidi generati dall'attivazione del sistema nervoso simpatico sotto stress, modellabili come semi-onde sinusoidali sovrapposte al segnale di base [36]. Il modello attuale non opera questa distinzione: se il paziente ha semplicemente caldo, l'EDA sale e il sistema potrebbe scambiare quella salita termica per una rampa pre-crisi, generando un falso allarme.

Per risolvere questo problema, prima di passare il dato alla Change Point Analysis il segnale grezzo dovrà essere processato tramite algoritmi di ottimizzazione convessa, come la *cvxEDA* proposta da Greco et al. [28], che separano matematicamente la com-

ponente legata al sudore termico da quella puramente nervosa. Il modello lavorerà e calcolerà il change point esclusivamente sulla componente fasica, aumentando la specificità del rilevamento delle anomalie. Per i segnali di HRV e ACC, dovranno essere implementati filtri passa-basso per rimuovere i picchi non fisiologici (come gli urti accidentali del braccio) e algoritmi di interpolazione per colmare le lacune di segnale (NaN) che si generano quando il sensore ottico perde temporaneamente l'aderenza cutanea [37].

7.9.2 Dalle soglie globali alle finestre temporali mobili

L'attuale validazione clinica si basa sulla funzione MATLAB `prctile()`, che calcola l'85° o il 75° percentile considerando l'intera matrice dei dati. Si tratta quindi di un approccio offline, che richiede la disponibilità dell'intera sessione prima dell'elaborazione. In un dispositivo indossabile che opera in tempo reale, invece, la baseline del paziente varia continuamente nel corso della giornata e il sistema non può conoscere in anticipo i valori futuri del segnale. La letteratura ha dimostrato che alterazioni della variabilità della frequenza cardiaca precedono l'insorgenza della crisi già nella fase pre-ittale [38], rendendo indispensabile un sistema di soglie dinamiche che catturi questi scostamenti in tempo reale.

Le soglie statiche dovranno pertanto essere sostituite da finestre mobili (*sliding windows*). Il sistema manterrà un buffer con solo gli ultimi 15 o 30 minuti di registrazione e ricalcolerà le regole di rilevamento su questo intervallo aggiornato. La domanda che il sistema si pone non sarà più se il valore di HRV è sotto la mediana globale della giornata, ma se è sotto la mediana degli ultimi 20 minuti. In questo modo il concetto di normalità viene continuamente aggiornato in base alle attività quotidiane del soggetto, riducendo i falsi allarmi dovuti alle normali fluttuazioni fisiologiche.

7.9.3 Standardizzazione intra-soggetto per il Random Forest

Il Random Forest attuale è stato addestrato per riconoscere il pericolo in base a valori assoluti delle feature fisiologiche. Tuttavia, ogni persona con ASD presenta una reattivi-

vità cutanea differente [37]: un soggetto ipo-reattivo potrebbe non superare mai i valori soglia appresi dal modello su dati di un soggetto iper-reattivo, rendendo il classificatore inefficace.

Per rendere il modello universalmente applicabile, le feature dovranno essere standardizzate tramite z-score prima di essere passate alla funzione `predict`. Applicando la trasformazione

$$z = \frac{x - \mu_w}{\sigma_w} \quad (7.1)$$

dove x è il valore corrente della feature, μ_w è la media e σ_w è la deviazione standard calcolate esclusivamente sulla finestra temporale mobile corrente w , l'algoritmo non leggerà più il valore assoluto del segnale ma l'intensità dello scostamento relativo rispetto alla baseline recente del soggetto. In questo modo, lo stesso modello potrà operare indistintamente su pazienti con reattività cutanea molto diversa, eliminando la dipendenza dai valori di grandezza assoluta su cui il classificatore è stato addestrato.

7.9.4 Transizione dall'analisi offline all'elaborazione in streaming

Nel codice sviluppato per questa tesi, l'analisi è di tipo offline: lo script carica l'intera matrice dei dati e la elabora dall'inizio alla fine, avendo a disposizione il segnale in tutta la sua durata. Un dispositivo predittivo indossabile non può permettersi questo approccio: deve prendere decisioni istante per istante, man mano che la fisiologia del paziente evolve. Questa transizione richiede di modificare l'architettura base del software, passando da uno script sequenziale a un sistema di elaborazione in streaming, attraverso i seguenti adeguamenti strutturali.

Architettura event-driven. Lo script non sarà più un blocco di comandi con inizio e fine, ma diventerà un processo perennemente attivo in background, basato su un sistema a *callback*: il codice rimane inattivo per preservare la batteria del dispositivo e si attiva automaticamente solo quando il bluetooth del braccialetto Empatica trasmette un nuovo pacchetto di dati.

Buffer circolare. Poiché non è computazionalmente utile conservare e rianalizzare continuamente ore di registrazione, il sistema utilizzerà una struttura dati a buffer circolare (coda FIFO: *First In, First Out*), dimensionata per contenere esclusivamente gli ultimi 15 minuti di dati. Ad ogni nuovo campione trasmesso dal sensore, questo viene inserito nel buffer mentre il campione più vecchio viene rimosso, mantenendo la dimensione della struttura costante nel tempo e garantendo che l'analisi operi sempre e soltanto sulla porzione di segnale più recente e clinicamente rilevante.

Esecuzione mirata della Change Point Analysis. Applicando la CPA esclusivamente al contenuto del buffer circolare, il costo computazionale si riduce drasticamente, rendendo l'algoritmo compatibile con i processori a risorse limitate di smartphone o smartwatch [31]. L'integrazione della Change Point Detection in un'architettura a cascata ha dimostrato di ridurre significativamente i falsi positivi rispetto all'uso del solo classificatore su serie temporali [31, 32]. L'obiettivo della CPA applicata al buffer è preciso: rilevare se, negli ultimi 15 minuti di acquisizione, il segnale elettrodermico ha subito una variazione brusca e statisticamente significativa della propria tendenza.

Il trigger di attivazione del Random Forest. Quando la CPA individua una variazione statisticamente significativa all'interno del buffer, il sistema non procede immediatamente alla classificazione, ma esegue prima una validazione incrociata dei segnali: verifica se, contestualmente allo sbalzo dell'EDA, siano presenti anche una riduzione dell'attività motoria e un calo della variabilità cardiaca. Solo se entrambe le condizioni risultano soddisfatte, le feature vengono fornite in ingresso al Random Forest per la classificazione finale e l'eventuale notifica al caregiver. Questa logica a due stadi consente di ridurre sensibilmente il numero di falsi allarmi e di limitare il ricorso al classificatore ai soli istanti fisiologicamente rilevanti, contenendo il costo computazionale complessivo del sistema. La capacità di operare in questo modo su un flusso di dati continuo rappresenta il presupposto fondamentale per la trasferibilità clinica del modello, garantendo quel lead time di 2–3 minuti necessario per intervenire tempestivamente prima dell'escalation della crisi [14].

7.9.5 Transfer learning: integrare la simulazione con i dati clinici reali

L'addestramento su dati sintetici è stato essenziale per compensare l'assenza di crisi nei dataset iniziali, e studi recenti hanno dimostrato che questo approccio può migliorare le performance dei classificatori fino all'11.9–15.5% rispetto all'addestramento su dati reali insufficienti [33]. Tuttavia, esso introduce inevitabilmente una differenza rispetto al profilo di segnale acquisito dai biosensori Empatica E4 in condizioni reali: il modello ha appreso le regole del pericolo fisiologico, ma non conosce ancora le caratteristiche specifiche del sensore da polso né il rumore di fondo che lo caratterizza [35]. Tra i metodi di data augmentation analizzati in letteratura, quelli di generazione parametrica risultano tra i più efficaci applicati a dataset di segnali fisiologici analoghi al nostro scenario [26], confermando la validità dell'approccio adottato in questo lavoro.

Non sarà tuttavia necessario ripartire da zero. Non appena saranno disponibili registrazioni cliniche contenenti reali episodi di aggressività, queste verranno etichettate manualmente e integrate nella matrice di addestramento sintetica (`T_train`). Il classificatore verrà quindi riaddestrato su questo dataset tramite `fitcensemble`: i dati artificiali continueranno a garantire il bilanciamento delle classi, fornendo al modello un numero sufficiente di esempi di pre-crisi, mentre la parte di dati reali permetterà al classificatore di adattarsi progressivamente alle caratteristiche elettriche specifiche del sensore. Questa strategia, già validata in letteratura nell'ambito della data augmentation per serie temporali fisiologiche [34, 26], permette di superare la difficoltà concreta di raccogliere un dataset reale completo e bilanciato in popolazioni cliniche rare. Il modello potrà così essere affinato progressivamente, integrando i nuovi dati clinici non appena questi si renderanno disponibili [33].

8. Generalizzazione del modello ibrido ai piloti aerei

8.1 Basi fisiologiche: ASD vs piloti

Il punto di partenza di questa generalizzazione è stato capire in che cosa la risposta fisiologica allo stress in un pilota sano differisca dalla disregolazione emotiva acuta tipica di un soggetto con ASD. Non si tratta di una distinzione puramente teorica: determina concretamente come vanno calibrate le soglie, di quanto anticipare l'etichetta e quali segnali fisiologici includere nel modello.

Dinamica temporale e carico allostatico. Uno degli errori metodologici più diffusi nella letteratura sullo stress è trattare la risposta fisiologica come un evento puntuale. Rosenblum et al. [10] mostrano invece che, analizzando i segnali su finestre temporali successive, lo stress cognitivo si manifesta come un carico che si accumula progressivamente (carico allostatico), sotto forma di plateau sostenuto piuttosto che di picco esplosivo. Questa evidenza ha una conseguenza diretta per il modello: il classificatore non deve cercare un'impennata acuta, ma imparare a riconoscere una rampa lenta che può durare 10–15 minuti [2]. La finestra di label shifting è stata di conseguenza estesa da 6 a 30 campioni rispetto al modello ASD, come si discuterà nella sezione 8.4.1.

EDA, modulazione vagale e frequenza cardiaca. A differenza dei soggetti ASD, i piloti presentano una risposta autonoma più omogenea e contenuta. Con un'EDA basale di soli $0.5\text{--}2\ \mu\text{S}$, le soglie di rilevamento devono essere calibrate sulla distribuzione individuale piuttosto che su valori normativi di popolazione. A tal proposito, Coelli et al. [11] dimostrano che l'ampiezza fasica del segnale elettrodermico subisce un'impennata netta sotto carico mentale, confermandola come marcatore primario dell'attivazione simpatica anche in questa popolazione. Sul fronte cardiaco, il dato più rilevante per la progettazione del modello viene da Wang et al. [2]: la frequenza cardiaca è il parametro più reattivo allo stress nei piloti, capace di registrare

l'attivazione simpatica già nei primissimi minuti del fattore di stress e di anticipare il successivo calo dell'HRV. Questa osservazione ha portato all'integrazione dell'HR come quinta feature, assente nel modello ASD.

Ruolo dell'accelerometro nel cockpit. Nel modello ASD, l'accelerometro registrava i micromovimenti legati allo stimming o ai blocchi comportamentali. Nel cockpit la situazione è opposta: il corpo del pilota è ancorato al sedile e le variazioni dell'ACC sono minime. Tuttavia, Yu et al. [12] mostrano che le accelerazioni traslazionali passive, inevitabili durante turbolenze, virate e variazioni di quota, agiscono come fattori di stress fisici diretti sul sistema nervoso autonomo, con una risposta EDA rilevabile in finestre di pochi secondi. Il segnale ACC è stato quindi mantenuto nel modello sia come marcatore della riduzione progressiva dell'attività motoria tipica del pilota sotto sforzo cognitivo elevato, sia per preservare l'informazione legata alle manovre di volo.

8.2 Giustificazione dell'approccio sintetico

La scelta di lavorare con un dataset interamente sintetico nasce da una necessità pratica: i dati fisiologici acquisiti su piloti in condizioni operative reali sono quasi sempre inaccessibili, vincolati da restrizioni di privacy, protocolli di sicurezza stringenti e costi di acquisizione molto elevati. Non si tratta di un limite specifico di questo lavoro, ma di una condizione comune a tutto il settore, tanto che anche in letteratura le fasi di proof-of-concept ricorrono frequentemente a simulatori certificati o a dati generati sinteticamente [2].

I parametri della funzione `genera_sessione_pilota_avanzata` non sono stati scelti arbitrariamente, ma derivano direttamente dai valori riportati in letteratura: la rampa EDA di 10–15 minuti, la riduzione dell'HRV del 25%, il picco di HR fino a ~ 110 bpm e il micro-congelamento dell'ACC riproducono dinamiche osservate sperimentalmente e documentate in studi recenti [10, 11]. Un secondo vantaggio riguarda la fase di validazione: disporre del controllo completo sul ground truth ha reso possibile misurare con precisione il lead time e verificare la separabilità delle classi in condizioni

note, un tipo di analisi che su dati osservazionali reali presupporrebbe un'annotazione manuale da parte di personale esperto, difficile da ottenere in questo contesto.

Tuttavia, i valori di performance ottenuti rappresentano una stima ottimistica, perché il dato sintetico è per costruzione privo del rumore che caratterizza un contesto operativo reale, ovvero artefatti da movimento, variazioni di temperatura e umidità cutanea, interferenze elettromagnetiche del cockpit. La validazione su dati reali acquisiti in ambiente di volo è quindi il passo successivo naturale di questo lavoro.

8.3 Generazione dei dati sintetici

Il dataset comprende $N = 10$ soggetti, ciascuno con 1000 campioni (500 minuti totali a 1 campione ogni 30 s), organizzati in una matrice a 5 variabili: EDA, EDA Slope, HRV RMSSD, ACC e HR. La dinamica è strutturata in tre fasi.

- **Pre-stress (rampa):** L'accumulo del carico cognitivo è simulato con un incremento lineare dell'EDA per 10–15 minuti. Contemporaneamente, l'HR sale rapidamente fino a ~ 110 bpm, riflettendo l'attivazione simpatica immediata documentata da Wang et al. [2].
- **Plateau:** HRV e ACC vengono moltiplicati per fattori riduttivi (rispettivamente fino al 75% e al 92%), riproducendo la progressiva perdita di variabilità cardiaca e la riduzione dell'attività motoria osservabili in un pilota sotto carico cognitivo sostenuto.
- **Recovery esponenziale:** Rosenblum et al. [10] documentano che al termine del fattore di stress i parametri fisiologici non tornano alla baseline in modo simultaneo. Per riprodurre questa asimmetria, la fase post-plateau è modellata con un decadimento esponenziale differenziato: l'HR rientra rapidamente ai valori di riposo, mentre l'EDA decade con una costante di tempo più lenta, seguendo il graduale riassorbimento della risposta simpatica cutanea.

L'istante di crisi è campionato casualmente nell'intervallo [700, 880] campioni e il numero di campioni di anticipo è estratto uniformemente da [20, 30], impedendo all'al-

goritmo di memorizzare la posizione temporale dell'evento e prevenendo l'overfitting temporale.

8.4 Architettura ibrida a due stadi

8.4.1 Label shifting e definizione del target

Per consentire la predizione anticipata, è stata applicata la tecnica del label shifting: l'etichetta "stress" non è assegnata all'istante del picco fisiologico, ma alla finestra temporale che lo precede. Nel modello ASD questa finestra corrispondeva a 6 campioni (3 minuti), calibrata sui valori di riferimento di Imbiriba et al. [14] per la predizione delle crisi comportamentali in soggetti con ASD. Nel presente modello è stata estesa a 30 campioni (15 minuti), perché in aviazione la correzione di un errore cognitivo o la modifica di una procedura richiedono tempi superiori a quelli dell'intervento di un caregiver.

8.4.2 Stadio 1: Change Point Analysis con baseline personalizzate

La CPA individua il punto di discontinuità strutturale del segnale EDA tramite `findchangepts`, con statistica lineare applicata sull'EDA smussata mediante media mobile a 5 campioni. Per ridurre i falsi positivi, il rilevamento è sottoposto a una validazione multimodale: l'evento viene confermato solo se l'EDA supera l'85° percentile della propria baseline individuale e se almeno una delle seguenti condizioni è contemporaneamente verificata:

- EDA Slope supera il 75° percentile individuale (velocità di incremento dell'attività simpatica).
- HRV RMSSD scende sotto l'85% della mediana individuale (ritiro vagale).

Richiedendo la concordanza di almeno due segnali indipendenti, questa struttura logica AND+OR limita i falsi positivi senza sacrificare la sensibilità, in linea con l'approccio multimodale raccomandato per i contesti safety-critical [11, 12]. Le soglie

vengono stimate sulla prima metà della sessione, che in uno scenario operativo reale corrisponderebbe alla finestra di calibrazione pre-volo. Il metodo ha prodotto una validazione positiva su tutti e 10 i soggetti analizzati.

8.4.3 Stadio 2: dal Random Forest al RUSBoost

Nel modello ASD, la classificazione era affidata a un Random Forest basato su Bagging. In presenza di un forte sbilanciamento delle classi, un classificatore di questo tipo tende a massimizzare l'accuratezza globale privilegiando sistematicamente la classe maggioritaria e annullando la sensibilità [39]. Apostol et al. [31] hanno mostrato come l'integrazione di un rilevatore di change point all'interno di una pipeline di classificazione consenta di contenere sistematicamente i falsi positivi rispetto all'impiego del solo classificatore su serie temporali multivariate; tale principio risulta direttamente trasferibile al contesto del presente lavoro. Nel contesto aeronautico simulato il rapporto effettivo tra le classi è di 32:1 a favore della calma, rendendo questo limite particolarmente critico.

Per affrontare questa criticità, si è optato per la transizione a RUSBoost (Random Under-Sampling Boosting), un algoritmo ad ensemble che unisce due tecniche complementari.

- **Random Under-Sampling:** Prima di ogni ciclo di apprendimento, la classe maggioritaria viene sottocampionata stocasticamente fino a raggiungere un rapporto di bilanciamento imposto a 3:1.
- **Boosting sequenziale (AdaBoost):** I classificatori deboli sono addestrati in sequenza; ad ogni iterazione i pesi dei campioni classificati erroneamente vengono aumentati, forzando l'albero successivo a specializzarsi sui casi più difficili [39].

Questa architettura massimizza la sensibilità per costruzione, minimizzando il rischio dei falsi negativi in contesti safety-critical [2].

8.4.4 Masking della recovery e filtro causale real-time

Masking della recovery. Come evidenziato da Rosenblum et al. [10], al termine di un evento stressante il corpo impiega molti minuti per riassorbire il sudore e ripristinare il tono vagale. Includere questa fase nel training etichettandola come calma introdurrebbe un data leakage: il modello imparerebbe ad associare un'EDA ancora elevata a uno stato di tranquillità, compromettendone la capacità predittiva. Per evitare questo errore, tutti i dati post-crisi vengono esclusi dal set di training tramite una maschera logica, in modo che il RUSBoost apprenda esclusivamente dalla fisiologia della transizione dalla calma alla fase di pre-stress.

Filtro causale asimmetrico. Lo score grezzo prodotto dal RUSBoost viene regolarizzato tramite una media mobile causale con parametri [14 0]: il sistema calcola la media degli score degli ultimi 14 campioni acquisiti, corrispondenti a 7 minuti, senza fare ricorso ad alcun campione futuro. Questa scelta garantisce la piena applicabilità real-time in cabina di pilotaggio, dove i dati futuri non sono disponibili per definizione [2]. La soglia di attivazione dell'allarme è fissata a zero: valori negativi indicano la prevalenza di uno stato di calma, valori positivi la prevalenza di uno stato di stress.

8.5 Protocollo di validazione: Leave-One-Out Cross-Validation

Per valutare la generalizzazione a piloti ignoti, è stata implementata una Leave-One-Out Cross-Validation (LOO-CV) su 10 fold: ad ogni iterazione, il modello viene addestrato sui dati di 9 piloti e testato sul decimo.

Tabella 8.1: Risultati della LOO-CV sul modello aeronautico ($N = 10$ piloti).

Metrica	Valore
AUC (Area Under Curve)	0.89 ± 0.04
Sensibilità (Recall)	79.0%
Specificità	99.1%
Lead Time Medio	10.1 min

Nota metodologica. La LOO-CV quantifica la robustezza inter-soggetto rispetto alla variabilità stocastica introdotta dal seed di generazione, non rispetto alla variabilità

fisiologica inter-individuale reale. I risultati forniscono pertanto una stima del limite superiore delle prestazioni attese su dati operativi: in uno scenario reale, il rumore fisiologico del contesto di volo tenderebbe a ridurre sia la sensibilità che l'AUC.

L'elevata specificità del 99.1% è in parte una conseguenza diretta dello sbilanciamento del dataset: con una distribuzione così asimmetrica, qualunque classificatore tenderebbe a privilegiare la classe maggioritaria. Il masking e il filtro causale contribuiscono a ridurre i falsi positivi durante le fasi di volo stabile prolungate, ma il valore assoluto ottenuto va comunque interpretato tenendo conto dello sbilanciamento tra le classi.

A titolo indicativo, l'AUC di 0.89 risulta compatibile con i valori in letteratura ottenuti su dati reali da dispositivi indossabili: Dalmeida & Masala (2021) riportano infatti $AUC \sim 0.83-0.85$ su ECG da smartwatch [25], pur tenendo presente che dataset di tipo sintetico tendono a produrre stime ottimistiche delle prestazioni reali.

Il lead time medio di 10.1 minuti soddisfa il requisito operativo per l'intervento nel cockpit [2]: tale preavviso consente di attivare procedure di mitigazione prima che la situazione diventi critica, tra cui il trasferimento dei comandi al co-pilota, l'inserimento del pilota automatico o la redistribuzione del carico cognitivo in anticipo rispetto alle manovre più impegnative.

8.6 Analisi dei risultati

8.6.1 Segnali fisiologici e predizione real-time

La figura 8.1 mostra la dinamica dei segnali sul fold con AUC massimo. Si osservano le quattro fasi: una baseline stabile, una rampa con $HR\uparrow/HRV\downarrow$, un plateau sostenuto e un recupero esponenziale post-crisi. Il classification score RUSBoost supera la soglia zero con un anticipo medio di 10.1 minuti, confermando la capacità del sistema di anticipare l'evento di stress..

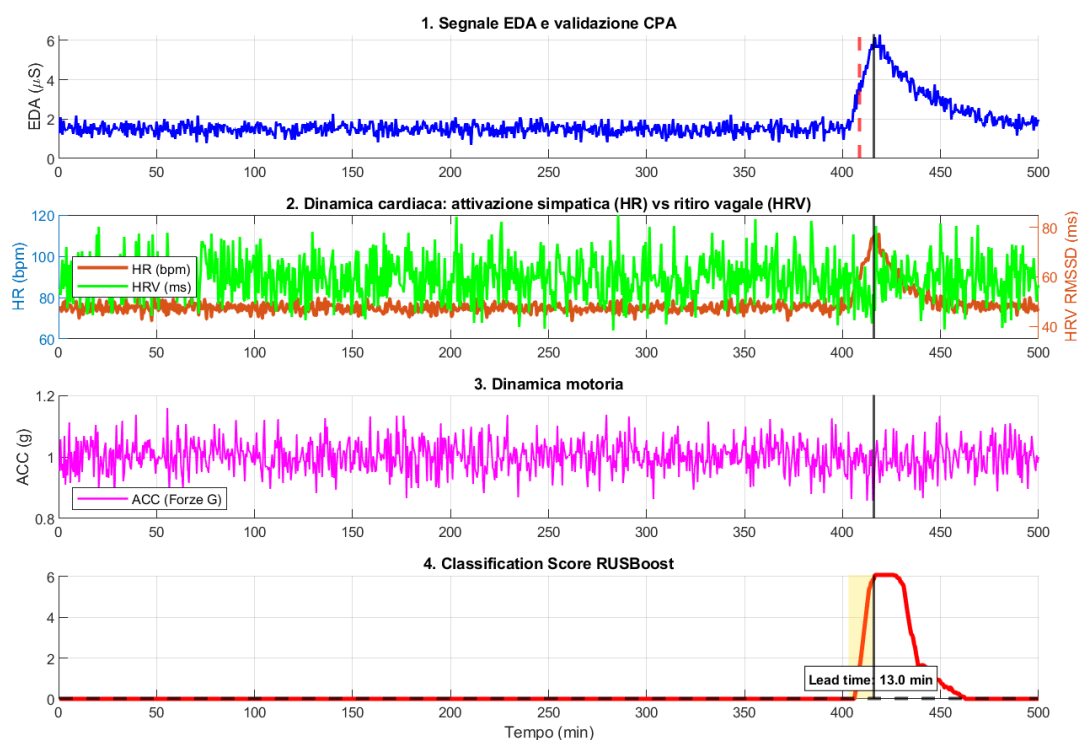


Figura 8.1: Dinamica rampa-plateau-esponenziale e predizione real-time. (1) Segnale EDA e validazione CPA. (2) Dinamica cardiaca: attivazione simpatica rapida (HR, asse sinistro) e ritiro vagale latente (HRV, asse destro). (3) Dinamica motoria: ACC piatto, conferma dell'immobilità nel cockpit. (4) Classification score RUSBoost con filtro causale e lead time.

8.6.2 Matrici di confusione e curva ROC

La Figura 8.2 mostra le matrici di confusione sul set di training e sul set di test ottenute dalla LOO-CV. La sensibilità del 79.0% e la specificità del 99.1% si mantengono stabili sui dati ignoti, indicando una buona capacità di generalizzazione a piloti mai visti durante l'addestramento. L'assenza di degrado significativo tra training e test esclude fenomeni di overfitting.

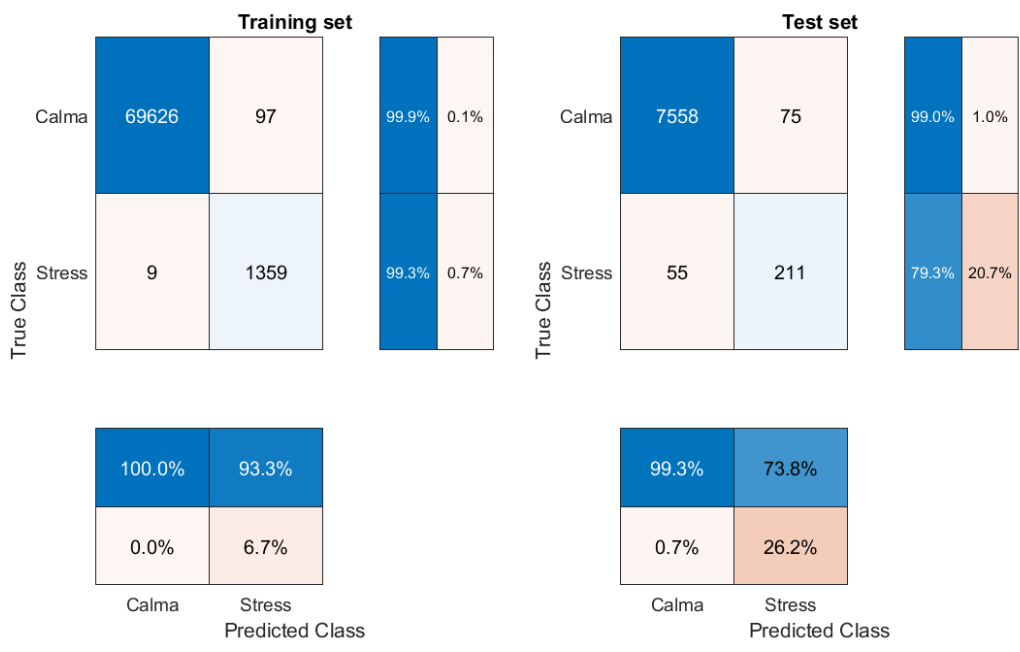


Figura 8.2: Matrici di confusione sul training set e sul test set.

La curva ROC in figura 8.3 mostra un'AUC di 0.89, superiore al riferimento di 0.84 riportato da Goodwin et al. [3] per la predizione delle crisi ASD e compatibile con i valori standard aeronautici di Dalmeida & Masala (AUC ~0.83–0.85) [25].

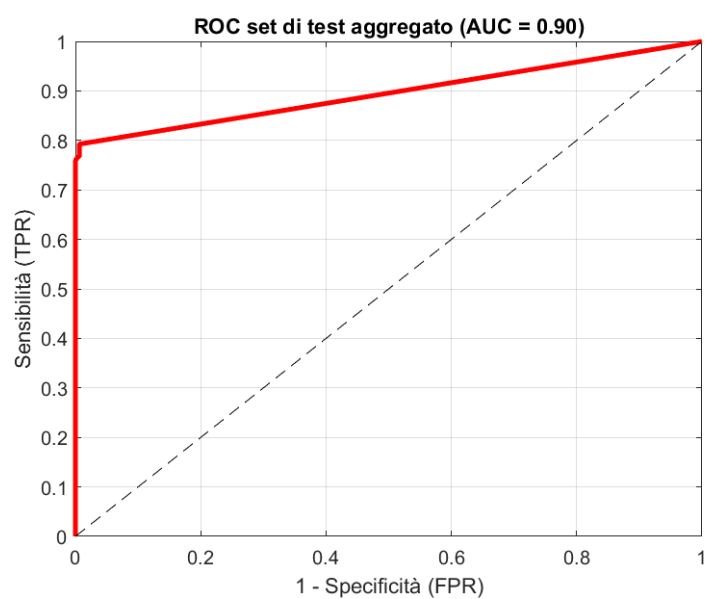


Figura 8.3: Curva ROC sul set di test (AUC = 0.89).

8.6.3 Importanza relativa delle feature

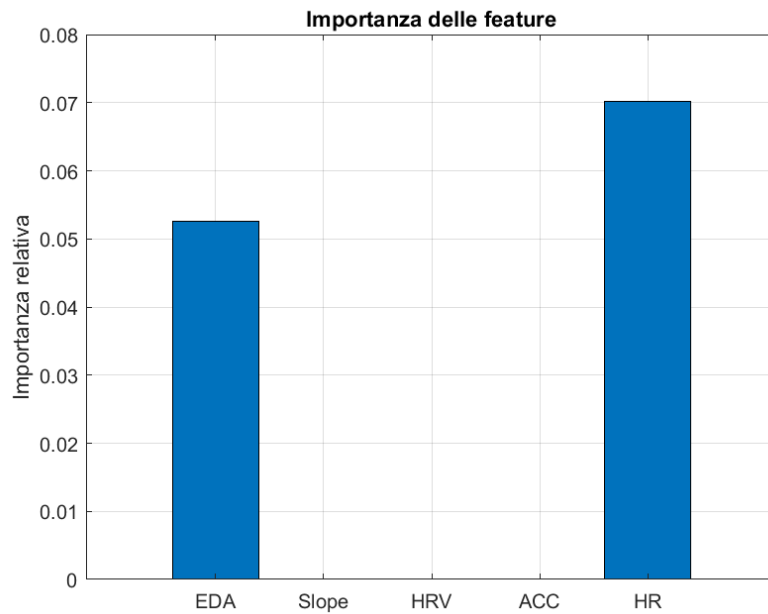


Figura 8.4: Importanza predittiva relativa delle 5 feature (media LOO-CV). L'HR emerge come predittore primario nel dataset sintetico, in linea con le evidenze di Wang et al. [2].

L'analisi del peso predittivo delle feature mostra che l'HR supera l'EDA come contributo principale alla classificazione. Il risultato è coerente con quanto riportato da Wang et al. [2]: la rampa lineare dell'HR nei dati sintetici è tuttavia priva del rumore fisiologico reale. In un contesto operativo, variabili come lo sforzo fisico o le variazioni di temperatura cutanea ne ridurrebbero l'affidabilità se usato come unico predittore, rendendo l'approccio a 5 feature necessario per una classificazione robusta.

8.6.4 Consistenza del modello e robustezza inter-soggetto

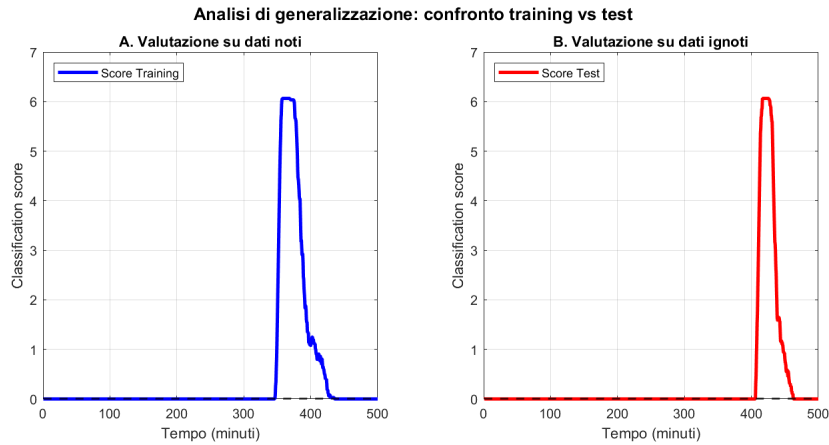


Figura 8.5: Verifica di consistenza: confronto diretto tra score su dati noti (training) e dati ignoti (test).

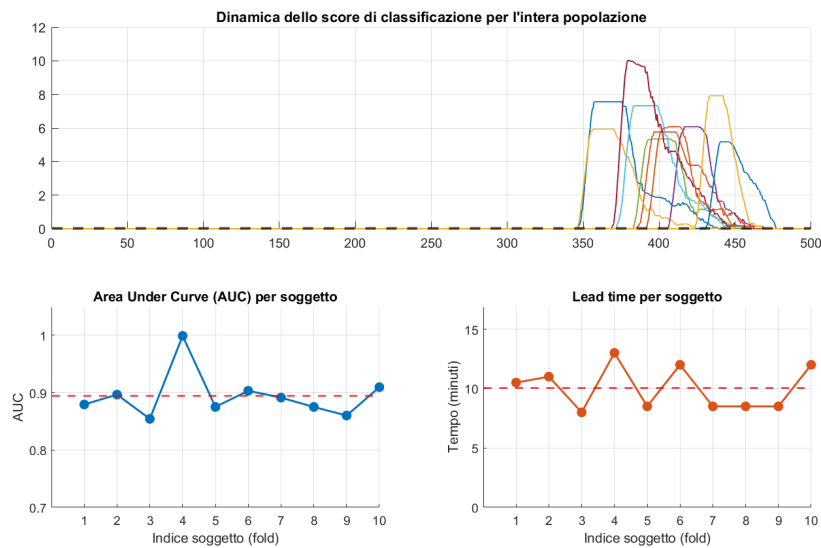


Figura 8.6: Robustezza statistica: fluttuazione di AUC e lead time sui 10 fold della LOO-CV. La varianza contenuta (± 0.04) indica una generalizzazione all'eterogeneità stocastica inter-soggetto.

Il confronto diretto tra le predizioni su soggetti di training e soggetti di test mostra andamenti di score coerenti tra le due condizioni, confermando l'assenza di overfitting. La fluttuazione di AUC tra i 10 fold (0.89 ± 0.04) indica una varianza inter-soggetto contenuta e un comportamento di generalizzazione stabile [2].

8.7 Confronto metodologico: modello ASD vs modello piloti

Tra il modello ASD e il modello aeronautico le differenze non sono solo numeriche. Oltre alla ricalibrazione dei parametri fisiologici, l’adattamento ha richiesto scelte architetturali sostanziali, riassunte nella tabella 8.2.

Tabella 8.2: Confronto metodologico tra modello ASD e modello piloti.

Parametro	Modello ASD	Modello piloti
Sensor fusion	4 feature (no HR)	5 feature (HR integrata)
Classificatore	Random Forest (Bagging)	RUSBoost
Validazione	Test statico ($N = 2$)	LOO-CV ($N = 10$)
Data leakage	vettore completo	masking logico post-crisi
Causalità	Filtri simmetrici	Media mobile
Label Shifting	6 campioni (3 min)	30 campioni (15 min)
Onset	Fisso	Stocastico [700–880]
Soglia decisionale	Soglia empirica 25%	Classification score = 0

La scelta di sostituire il Random Forest con RUSBoost non è casuale: rappresenta una risposta diretta allo sbilanciamento 32:1 del dataset aeronautico, più pronunciato rispetto al caso ASD. Il masking della recovery e il filtro causale [14 0] rappresentano adattamenti specifici per l’operatività in tempo reale, assenti nel modello ASD in cui la validazione veniva condotta offline. La finestra di label shifting estesa a 15 minuti riflette invece la natura graduale del MWL aeronautico, in contrasto con il meltdown autistico, evento acuto che decade in circa 3 minuti.

8.8 Limiti e prospettive

Il principale limite di questo lavoro è la natura sintetica del dataset, discussa in dettaglio nella sezione 8.2. La validazione su dati reali, preferibilmente acquisiti su piloti equipaggiati con dispositivi indossabili come l’Empatica E4 o equivalenti validati [25], rappresenta il naturale passo successivo.

Un ulteriore aspetto riguarda la frequenza respiratoria e il suo legame con l’HRV tramite la RSA, che Coelli et al. [11] identificano come uno dei predittori più potenti

dello stress cognitivo nei piloti. In questo lavoro non è stata inclusa tra le feature del modello principalmente perché i dati utilizzati sono sintetici: aggiungere una componente respiratoria simulata avrebbe richiesto di modellare esplicitamente come il respiro si propaga sul segnale HRV, introducendo ipotesi aggiuntive difficili da calibrare senza dati reali di riferimento. Va inoltre considerato che Wang et al. [2] e Dalmeida & Masala [25], che costituiscono i principali riferimenti di questo lavoro, ottengono risultati comparabili (AUC 0.83–0.85) usando solo segnali da polso e senza correzione RSA. La frequenza respiratoria resta comunque una feature che vale la pena includere in una futura validazione su dati reali, dove permetterebbe di separare il contributo vagale puro da quello respiratorio nell'interpretazione dell'HRV.

Nonostante questi limiti, la generalizzazione ha dimostrato come l'architettura ibrida CPA+RUSBoost rappresenti uno schema metodologico solido e trasferibile tra popolazioni fisiologicamente diverse. Perché questo avvenga in modo efficace, è tuttavia necessario ridefinire soglie, label shifting e classificatore sulla base delle dinamiche specifiche del nuovo contesto applicativo.

9. Acquisizione sperimentale su soggetti sani tramite braccialetto Empatica EmbracePlus

9.1 Il dispositivo Empatica EmbracePlus

L'Empatica EmbracePlus è un braccialetto indossabile realizzato per la raccolta continua di segnali fisiologici in contesti di vita reale. Rispetto alla versione precedente E4, presenta un sistema sensoriale più completo: integra un sensore PPG con LED a tre lunghezze d'onda (rosso, infrarosso e verde) per la stima della frequenza cardiaca e della variabilità del battito periferico (PRV), un sensore EDA con elettrodi incorporati nel cinturino in silicone, un accelerometro e giroscopio triassiali per il monitoraggio del movimento, e un sensore di temperatura cutanea. Le frequenze di campionamento sono 64 Hz per il PPG e l'accelerometro, e 4 Hz per EDA e temperatura. È tuttavia fondamentale distinguere tra i dati grezzi acquisiti dal dispositivo e quelli effettivamente accessibili tramite la piattaforma CareLab. I dati grezzi, archiviati in formato Avro, contengono il segnale campionato alle frequenze sopra indicate. I file CSV esportati dal portale CareLab appartengono invece a un livello di elaborazione completamente diverso, denominato *digital biomarkers aggregated per minute*: ogni riga del CSV corrisponde a una finestra temporale di 60 secondi e contiene il risultato di un calcolo eseguito da Empatica sull'intera finestra, non un singolo campione grezzo del sensore.

Questo comporta implicazioni rilevanti per l'analisi:

- **EDA – SCL.** Il valore esportato non è il segnale EDA grezzo, ma la stima della componente tonica (SCL) calcolata sull'intera finestra di 60 secondi. Le Skin Conductance Responses fasiche (SCR), che sono i marcatori fisiologici più informativi dell'arousal simpatico acuto e hanno una durata tipica di 1–5 secondi, sono completamente invisibili a questa risoluzione temporale. Non è inoltre possibile applicare metodi di decomposizione tonica-fasica alternativi (co-

me cvxEDA) poiché il segnale grezzo a 4 Hz non è disponibile nel formato di esportazione standard.

- **PRV – RMSSD.** Il valore esportato è il RMSSD calcolato sulle differenze tra i picchi sistolici rilevati nel segnale PPG all'interno del minuto. Se la qualità del segnale PPG è insufficiente e vengono rilevati troppo pochi picchi validi, il campione viene marcato nel campo `missing_value_reason` e non compare nel CSV. Il PRV esportato non è quindi un segnale continuo, ma uno scatter di valori la cui densità dipende dal contatto del sensore.
- **Frequenza cardiaca (BPM) e temperatura.** Rappresentano rispettivamente la media della pulse rate e il valore medio di temperatura cutanea stimati nella finestra di 60 secondi, rendendoli insensibili a variazioni fisiologiche su scale temporali inferiori al minuto.
- **Accelerometria.** Il campo esportato è la deviazione standard della magnitudine dell'accelerometro triassiale nella finestra, non i campioni grezzi a 64 Hz.

Pertanto i dati analizzati hanno una risoluzione temporale di 60 secondi, indipendentemente dalla frequenza di campionamento del dispositivo, poiché le frequenze di campionamento originali si riferiscono all'acquisizione interna del sensore e non ai dati disponibili per l'analisi tramite l'esportazione standard di CareLab.

I dati vengono gestiti tramite la piattaforma CareLab di Empatica, composta da un'applicazione installabile su smartphone denominata CareLab App, che si connette al dispositivo via bluetooth e carica i dati sul cloud, e dal portale web CareLab Portal, dal quale il ricercatore può monitorare le sessioni ed esportare i segnali in formato CSV. Nell'esportazione è presente il campo `missing_value_reason`, che identifica i campioni non validi e consente di filtrarli in fase di pre-elaborazione. I dati grezzi sono archiviati in formato UTC, con timestamp espressi secondo lo standard ISO 8601. In fase di analisi, i timestamp sono stati convertiti al fuso orario locale Europe/Rome (CEST, UTC+2) tramite la funzione `datetime` di MATLAB. Gli orari riportati nei grafici e nel testo fanno pertanto riferimento all'ora locale italiana.

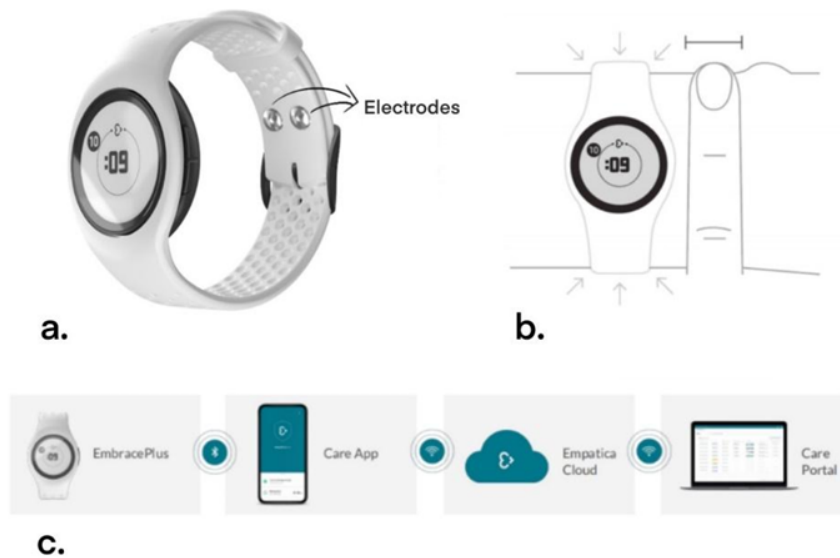


Figura 9.1: *Note.* a. Sensori EDA; b. Posizionamento corretto; c. Sincronizzazione dei dati. Immagini tratte da Empatica Inc. (2024).

9.1.1 Protocollo sperimentale

L'acquisizione si è svolta in un ambiente privo di distrazioni esterne significative. Il braccialetto è stato indossato sul polso non dominante di ciascun partecipante, posizionato a circa un dito dall'articolazione del polso e fissato con il cinturino in silicone con una pressione sufficiente a garantire il contatto dei sensori con la cute, ma senza stringere al punto da ostacolare la circolazione. Prima dell'indossaggio è stato applicato un gel conduttivo sugli elettrodi EDA, per ridurre l'impedenza cutanea e migliorare la qualità del segnale.

Il protocollo scelto è stato la visione di un film per tutta la durata della sessione, senza alcun compito aggiuntivo. Questa scelta è stata motivata dall'esigenza di limitare gli artefatti da movimento e di osservare la risposta fisiologica a uno stimolo emotivo naturale in un contesto ecologico. I due partecipanti hanno ricevuto stimoli di diversa natura emotiva:

- **Soggetto A:** film autobiografico, pensato per stimolare un coinvolgimento emotivo di tipo riflessivo. Sessione del 12 aprile 2026, ore 21:30–23:15.
- **Soggetto B:** film thriller, con stimoli di tensione, allerta e arousal acuto. Sessione del 10 aprile 2026, ore 14:30–15:30.

9.1.2 Partecipanti

Tabella 9.1: Caratteristiche dei partecipanti e dettagli delle sessioni di acquisizione.

	Soggetto A	Soggetto B
Sesso	Maschile	Femminile
Età	27 anni	26 anni
Stato di salute	Sano	Sana
Task	Film autobiografico	Film thriller
Data acquisizione	12 aprile 2026	10 aprile 2026
Orario sessione	21:30–23:15 (CEST)	14:30–15:30 (CEST)
ID CareLab	1-1-21	1-1-211311

9.1.3 Segnali acquisiti

Per entrambi i soggetti, i dati esportati da CareLab comprendono cinque segnali: frequenza cardiaca (bpm), variabilità periferica del battito PRV-RMSSD (ms), componente tonica dell'EDA o SCL (μS), temperatura cutanea ($^{\circ}\text{C}$) e attività motoria espressa come deviazione standard dell'accelerometro triassiale (g). Il segnale PRV è visualizzato come scatter plot in entrambe le figure poiché, nell'esportazione CareLab, il valore RMSSD non è il risultato di un campionamento continuo ma viene calcolato da Empatica sui picchi sistolici rilevati nel segnale PPG grezzo all'interno di ciascuna finestra di 60 secondi. Se in un determinato minuto la qualità del segnale PPG è insufficiente e il numero di picchi validi rilevati è troppo basso per il calcolo, il campione viene marcato nel campo `missing_value_reason` e non compare nel file CSV. La discontinuità del segnale PRV non è quindi una caratteristica intrinseca della metrica RMSSD, ma una conseguenza del fatto che si lavora su digital biomarkers aggregati per minuto anziché sul segnale grezzo: con accesso al BVP grezzo a 64 Hz sarebbe possibile calcolare il RMSSD su ogni coppia di battiti consecutivi, ottenendo una serie temporale densa. Nel formato di esportazione standard, la densità dei punti scatter dipende dalla qualità del contatto del sensore minuto per minuto.

Tutti i segnali analizzati presentano una risoluzione temporale di 60 secondi, corrispondente alla finestra di aggregazione dell'esportazione CareLab. Le frequenze di campionamento hardware (64 Hz per PPG e accelerometro, 4 Hz per EDA e tempera-

tura) si riferiscono all'acquisizione interna del dispositivo e non ai dati effettivamente analizzati, che sono il prodotto di elaborazioni algoritmiche eseguite da Empatica sull'intera finestra temporale.

9.1.4 Analisi dei segnali per soggetto

Soggetto B – Film thriller

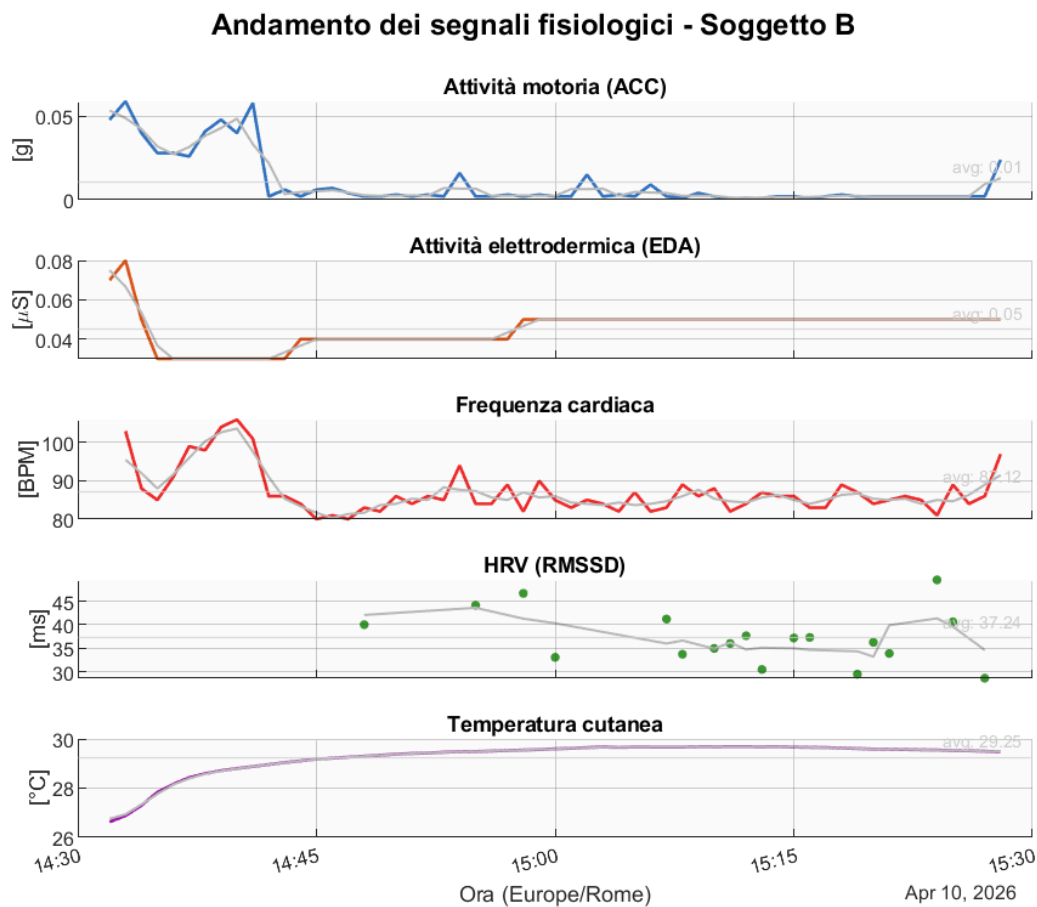


Figura 9.2: Segnali fisiologici acquisiti tramite Empatica EmbracePlus durante la visione di un film thriller (soggetto B).

La maggior parte dei segnali del soggetto B presentano una continuità accettabile, ad eccezione del PRV. Alcuni valori si discostano tuttavia dai range attesi per una donna sana di 26 anni, aspetto che verrà approfondito nel seguito.

Frequenza cardiaca. Il segnale mostra un valore iniziale di circa 100–105 BPM, che si riduce progressivamente fino a stabilizzarsi tra 80 e 85 BPM per il resto della

sessione. Il range fisiologico di riferimento per una donna di questa età in condizioni di riposo è generalmente compreso tra 60 e 80 BPM. Il valore elevato nei primi minuti è compatibile con una risposta di attivazione anticipatoria; il plateau successivo, pur rientrando nei limiti fisiologici, si colloca al limite superiore del range atteso per una sessione di visione passiva, e potrebbe riflettere un arousal sostenuto indotto dal film thriller o, in misura minore, artefatti da movimento residuo sul sensore PPG.

PRV-RMSSD. I valori si distribuiscono tra circa 29 ms e 47 ms. Per una donna sana di questa età, la letteratura riporta valori di RMSSD a riposo generalmente compresi tra 35 e 70 ms, con medie attorno a 40–50 ms. Il valore più basso registrato (~29 ms) scende al di sotto di questo range e potrebbe riflettere un momento di maggiore attivazione simpatica, ma la densità ridotta dei punti rende difficile trarre conclusioni solide. Non è inoltre possibile escludere che alcuni valori bassi siano in parte attribuibili a una qualità del segnale PPG non ottimale, con conseguenti errori nel calcolo degli intervalli inter-battito.

EDA – SCL. È il segnale più problematico dell'intera sessione. I valori registrati oscillano tra 0.04 e 0.08 μS , nell'ordine dei centesimi di microsiemens, mentre la letteratura colloca i valori tipici di SCL in soggetti sani a riposo tra 1 e 20 μS , con medie di 2–5 μS anche in condizioni di scarsa attivazione. Si tratta quindi di valori inferiori di almeno un ordine di grandezza rispetto al range atteso, un'anomalia difficilmente spiegabile con il solo compito passivo di visione di un film. L'ipotesi più plausibile è un contatto insufficiente tra gli elettrodi del cinturino e la cute, che il gel conduttivo non è riuscito a compensare adeguatamente.

Temperatura cutanea. Nel corso della sessione il segnale sale progressivamente da circa 26.5 °C a ~29 °C, senza mai stabilizzarsi. Per soggetti a riposo in ambiente controllato, la temperatura cutanea al polso è tipicamente compresa tra 31 e 35 °C. Il mancato raggiungimento del plateau suggerisce che il sensore non abbia mai raggiunto un equilibrio termico stabile con la cute, verosimilmente a causa di un indossaggio non sufficientemente aderente o di una temperatura ambientale particolarmente bassa

durante la sessione. Anche questi valori non possono quindi essere considerati una stima affidabile della temperatura cutanea reale.

Attività motoria (ACC). Dopo una breve fase di assestamento nei primi dieci minuti (~ 0.05 g), il segnale si porta su valori prossimi allo zero per tutta la durata della sessione, in linea con il protocollo di visione passiva. È l'unico segnale che si comporta come atteso.

Soggetto A – Film autobiografico

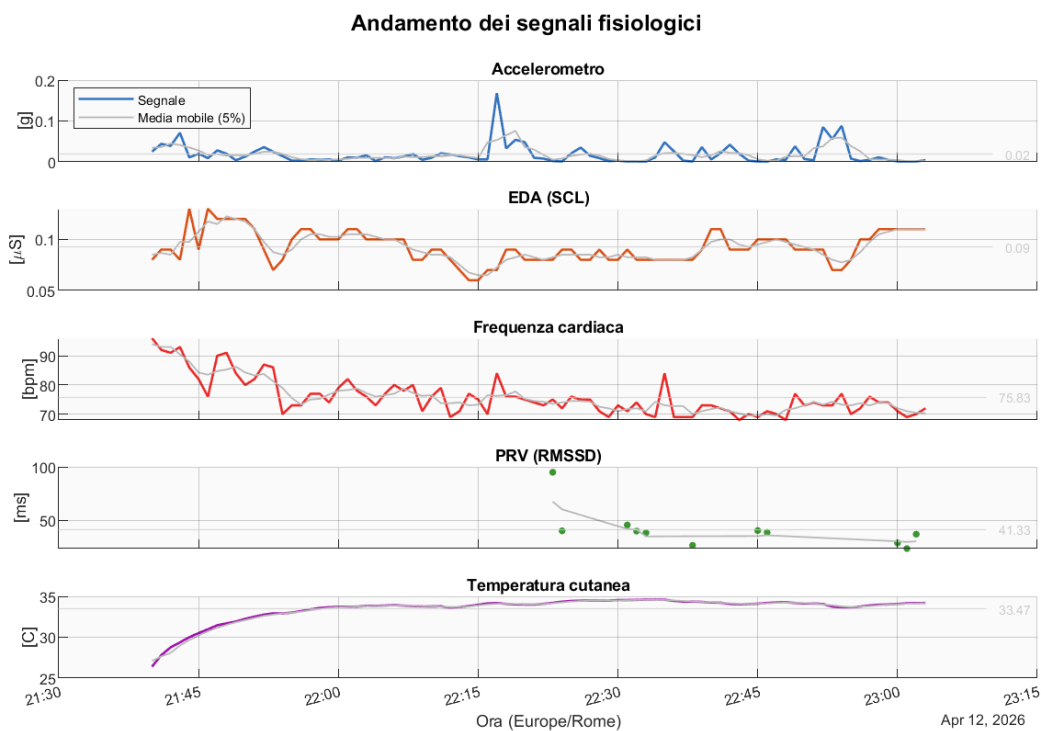


Figura 9.3: Segnali fisiologici acquisiti tramite Empatica EmbracePlus durante la visione di un film autobiografico (soggetto A).

Per il soggetto A le criticità di acquisizione sono più estese rispetto al soggetto B e riguardano più segnali contemporaneamente, rendendo difficile una lettura fisiologica affidabile dell'intera sessione.

Attività motoria (ACC). Il segnale presenta alcuni picchi distribuiti nel corso della sessione, tra cui uno particolarmente evidente (~ 0.17 g) intorno alle ore 22:18,

probabilmente riconducibile a un aggiustamento posturale. Per il resto i valori rimangono al di sotto di 0.05 g, compatibili con un soggetto seduto e sostanzialmente immobile. È il segnale più affidabile dell'intera registrazione.

EDA – SCL. I valori si collocano tra circa 0.06 e 0.13 μS , leggermente superiori rispetto al soggetto B ma comunque di almeno un ordine di grandezza al di sotto dei range attesi per un uomo sano di 27 anni, che in letteratura si collocano tipicamente tra 2 e 10 μS anche in condizioni di bassa attivazione emotiva. La variabilità temporale è maggiore, con alcuni aumenti nelle fasi iniziali e una riduzione intorno alle ore 22:05–22:10; tuttavia, data la qualità complessiva del segnale, non è possibile attribuire con certezza queste fluttuazioni a risposte emotive reali piuttosto che ad artefatti da movimento del polso o da variazioni del contatto degli elettrodi nel corso della lunga sessione.

Frequenza cardiaca. Il segnale parte da circa 90–92 BPM e scende progressivamente fino a stabilizzarsi attorno a 75 BPM, valore che rientra nel range fisiologico atteso per un uomo di 27 anni a riposo (60–80 BPM). La discesa iniziale è compatibile con il rilassamento progressivo durante la visione, e le oscillazioni tra 70 e 82 BPM nella fase centrale sono del tutto nella norma. È uno dei pochi segnali che mostra un andamento plausibile e coerente con i riferimenti bibliografici per questa fascia demografica.

PRV-RMSSD. Questo è il segnale più critico: nell'arco di quasi due ore sono presenti appena 10–12 punti scatter, il che rende il segnale inutilizzabile al fine di compiere analisi. Per un uomo sano di 27 anni i valori di RMSSD a riposo attesi si collocano tra 40 e 80 ms circa, con media intorno a 47 ms. Il valore massimo registrato (~ 95 ms) supera considerevolmente questo range, ed è probabilmente un artefatto da rilevamento errato degli intervalli inter-battito, coerente con un segnale PPG di qualità non ottimale. La media calcolata (41.33 ms) ricade per caso nell'intervallo atteso, ma essendo ottenuta da pochissimi punti non ha alcun valore statistico.

Temperatura cutanea. Il segnale sale da circa 26 °C fino a stabilizzarsi tra 33 e 34 °C intorno alle ore 22:00, dove rimane con lievi fluttuazioni fino al termine della sessione. A differenza del soggetto B, il soggetto A raggiunge un plateau all'interno del range fisiologico atteso (31–35 °C al polso a riposo), dopo una fase di salita iniziale di circa 30 minuti. Una volta raggiunta la stabilizzazione, questo è il segnale più affidabile tra quelli acquisiti per questo soggetto.

9.1.5 Problematiche di acquisizione

Entrambe le sessioni hanno presentato criticità che limitano l'affidabilità dei dati raccolti, con gravità diversa tra i due soggetti.

Il problema più evidente, comune a entrambi i partecipanti, riguarda il segnale EDA. I valori registrati si collocano nell'ordine dei centesimi di $\mu\text{Siemens}$ (0.04–0.13 μS), almeno un ordine di grandezza al di sotto dei range riportati in letteratura per soggetti sani a riposo (2–20 μS). Un'anomalia di questa entità non è spiegabile con il solo task passivo: indica con buona probabilità un contatto insufficiente tra gli elettrodi del cinturino in silicone e la cute, che nemmeno l'applicazione del gel conduttivo ha risolto in modo soddisfacente. In acquisizioni effettuate a livello del polso, fattori come la secchezza cutanea, la pressione di indossaggio e la temperatura ambientale possono tutti contribuire al degrado del contatto elettrodo. A questa criticità hardware si aggiunge una limitazione strutturale del formato di esportazione CareLab. Il valore EDA disponibile nel CSV non è il segnale grezzo a 4 Hz, ma la stima della componente tonica (SCL) calcolata algebricamente su una finestra di 60 secondi. Le Skin Conductance Responses fasiche (SCR), che costituiscono il marcatore fisiologico più informativo dell'arousal simpatico acuto e hanno una durata tipica di 1–5 secondi, sono completamente invisibili a questa risoluzione temporale, indipendentemente dalla qualità del contatto elettrodo. Anche con un contatto ottimale, il formato di esportazione standard renderebbe impossibile rilevare risposte emotive puntuali agli eventi del film e applicare metodi di decomposizione tonica-fasica alternativi come *cvxEDA*, che richiedono accesso al segnale grezzo. Le due criticità, ovvero qualità del contatto e risoluzione temporale, si sommano quindi, rendendo il segnale EDA espor-

tato non utilizzabile per l'analisi della risposta elettrodermica nella presente sessione sperimentale.

Per il soggetto B si aggiunge il problema della temperatura cutanea: il segnale non ha mai raggiunto un plateau nel corso dell'intera sessione, fermandosi a circa 29 °C anziché attestarsi nel range fisiologico atteso di 31–35 °C, il che indica che l'aderenza del sensore alla cute non era ottimale per tutta la durata della registrazione. Anche il PRV ha restituito un numero ridotto di campioni validi, rendendo impossibile qualsiasi analisi affidabile della variabilità autonoma.

Per il soggetto A le criticità sono ancora più estese: il segnale PRV ha prodotto appena 10–12 punti scatter in quasi due ore di sessione, rendendolo statisticamente inutilizzabile. Il segnale PPG sottostante, da cui derivano sia la frequenza cardiaca che il PRV, ha verosimilmente attraversato tratti prolungati di qualità insufficiente, come confermato dalla presenza di campioni marcati nel campo `missing_value_reason` dell'esportazione CareLab. La lunga durata della sessione può aver ulteriormente peggiorato nel tempo il contatto sensoriale, anche nei segnali che inizialmente sembravano stabili.

I dati raccolti con EmbracePlus in questo contesto sperimentale vanno quindi letti con la dovuta cautela, specialmente per quanto riguarda i segnali EDA e PRV. Per eventuali sessioni future, sarebbe opportuno verificare la qualità del segnale in tempo reale tramite la CareLab App prima di avviare la registrazione, attendere almeno 5 minuti dall'indossaggio per consentire al sensore di stabilizzarsi sulla cute, e limitare la durata complessiva della sessione per ridurre il rischio di degrado del contatto nel tempo.

10. Conclusioni

10.1 Sintesi del lavoro

Il presente lavoro ha sviluppato e validato un sistema di allerta precoce per la predizione di stati aggressivi legati alla disregolazione del Sistema Nervoso Autonomo, con applicazione primaria alla popolazione con ASD e con una generalizzazione al dominio aeronautico. Il filo conduttore metodologico è stato lo stesso lungo tutti i nove capitoli precedenti: partire dai segnali fisiologici acquisiti da un dispositivo indossabile da polso, costruire una rappresentazione individuale dello stato basale del soggetto, rilevarne le discontinuità strutturali tramite Change Point Analysis e classificarle tramite Random Forest addestrato con tecniche di Data Augmentation. A complemento del lavoro modellistico, è stata condotta una sessione sperimentale su due soggetti sani tramite il braccialetto Empatica EmbracePlus, con l'obiettivo di valutare la pipeline di acquisizione in un contesto controllato. L'analisi dei segnali esportati dalla piattaforma CareLab ha evidenziato due ordini di criticità distinti: da un lato, problemi hardware legati alla qualità del contatto degli elettrodi EDA e del sensore ottico PPG; dall'altro, una limitazione strutturale del formato di esportazione, che fornisce *digital biomarkers aggregati per minuto* anziché il segnale grezzo, rendendo invisibili le risposte fasiche elettrodermiche (SCR) e riducendo la densità dei campioni PRV disponibili. Queste evidenze, oltre a motivare il ricorso alla generazione sintetica dei dati di addestramento adottata nei capitoli precedenti, forniscono indicazioni operative concrete per il miglioramento del protocollo di acquisizione in sessioni future.

La pipeline completa integra cinque fasi distinte. Il pre-processing (capitolo 4) ha decomposto l'EDA nelle sue componenti tonica e fasica, preparato il segnale HRV dagli intervalli IBI estratti dal PPG e normalizzato l'accelerometria. L'estrazione delle feature (capitolo 5) ha costruito un dataset multimodale con finestre scorrevoli, calibrando ogni soglia sulla baseline individuale piuttosto che su valori normativi di popolazione, in risposta all'elevata eterogeneità intrasoggetto documentata nella letteratura sull'ASD. I modelli predittivi (capitolo 6) hanno introdotto la baseline dinamica e la CPA come approccio non supervisionato, integrandoli con un classificatore Random

Forest soggetto-specifico. Il modello ibrido (capitolo 7) ha combinato i due approcci in un'architettura a cascata, validata su dati sintetici con analisi di robustezza temporale e al rumore. La generalizzazione ai piloti (capitolo 8) ha dimostrato la trasferibilità del framework a un dominio fisiologicamente analogo ma strutturalmente diverso, con adattamenti architetturali sostanziali. Il capitolo 9 ha infine documentato un'acquisizione sperimentale su due soggetti sani tramite Empatica EmbracePlus, fornendo un riscontro empirico diretto sulle criticità della pipeline di raccolta dati e indicazioni operative per sessioni future.

10.2 Risultati principali

10.2.1 Modello ibrido ASD

Il modello ibrido CPA+RF ha raggiunto, su dati sintetici calibrati sulla letteratura, una sensibilità dell'85.7%, una specificità del 100% e un'AUC di 0.85, con un anticipo medio sull'evento di crisi di 2.5 minuti. Questi valori si collocano in linea con gli standard di riferimento: Goodwin et al. [3] riportano un'AUC di 0.84 con una finestra di predizione di 3 minuti, replicata da Imbiriba et al. [14] su un campione più ampio con AUC di 0.80.

L'analisi dell'importanza delle feature ha evidenziato che il predittore più informativo non è il valore assoluto dell'EDA, ma la sua derivata temporale (EDA Slope), che cattura la velocità di accumulo della risposta simpatica piuttosto che la sua ampiezza istantanea. Questo risultato ha una diretta implicazione clinica: il sistema rileva la direzione della traiettoria fisiologica, non il superamento di una soglia assoluta, rendendolo adattabile all'eterogeneità caratteristica della popolazione autistica.

L'analisi di robustezza temporale ha confermato la stabilità del modello al variare della posizione della crisi nella sessione, con AUC compresa tra 0.84 e 0.86 e lead time stabile nell'intervallo 2.4–2.6 minuti su tutti gli onset testati. Il test di stress a +6 dB ha invece evidenziato la vulnerabilità strutturale dell'EDA Slope al rumore, con un crollo della sensibilità dall'85.7% al 14.3%. L'intervento correttivo, ovvero pre-processing della slope tramite media mobile a 5 campioni e randomizzazione dell'ampiezza della

rampa sintetica, ha ripristinato completamente le prestazioni originali, portando l'AUC a 0.91 e il lead time a 2.5 minuti nel modello robusto.

10.2.2 Generalizzazione al dominio aeronautico

L'adattamento del modello al dominio aeronautico ha richiesto modifiche su più livelli, riassunte nella Tabella 8.2. La sostituzione del Random Forest con RUSBoost nasce dalla necessità di gestire uno sbilanciamento di classe 32:1, ben più accentuato di quello presente nel dataset ASD. L'estensione del label shifting da 6 a 30 campioni riflette la natura progressiva del MWL: a differenza della crisi ASD, che si manifesta come un picco acuto, il sovraccarico cognitivo del pilota si accumula in un plateau sostenuto di 10–15 minuti, e la finestra di etichettatura deve essere abbastanza ampia da catturarlo. L'aggiunta dell'HR come quinta feature copre l'attivazione simpatica rapida documentata da Wang et al. [2], che nel contesto aeronautico precede la riduzione dell'HRV con un anticipo significativo. Il masking della recovery e il filtro causale [14 0] garantiscono infine l'applicabilità in tempo reale e prevengono il data leakage durante la fase di addestramento. La validazione tramite LOO-CV su 10 soggetti sintetici ha prodotto AUC di 0.89 ± 0.04 , sensibilità del 79.0%, specificità del 99.1% e lead time medio di 10.1 minuti. La varianza contenuta tra i fold suggerisce che il modello generalizza in modo stabile, senza risentire in modo marcato dell'eterogeneità tra soggetti. Il valore di AUC si colloca al di sopra dei benchmark aeronautici su dati reali di Dalmeida & Masala [25] (AUC ~ 0.83 – 0.85), un risultato atteso considerando che le stime su dati sintetici tendono ad essere ottimistiche per costruzione.

10.2.3 Acquisizione sperimentale con Empatica Embrace-Plus

Le due sessioni condotte su soggetti sani durante la visione di film (capitolo 9) non hanno prodotto dati utilizzabili per la validazione del modello. Tuttavia, hanno fornito indicazioni pratiche molto utili sulla principale criticità dell'intera pipeline: la qualità del segnale EDA acquisito al polso. In entrambe le sessioni, i valori di SCL sono risultati nell'ordine dei centesimi di μS , quindi almeno un ordine di grandezza inferiori

rispetto ai range fisiologici attesi. Questo è probabilmente dovuto a un contatto non ottimale tra gli elettrodi del cinturino in silicone e la pelle. Anche il segnale PRV si è rivelato poco affidabile, con un numero di campioni validi troppo basso per consentire analisi statistiche robuste, soprattutto nel caso del soggetto con la sessione più lunga. Al contrario, la frequenza cardiaca ha mostrato un andamento coerente con i valori fisiologici per entrambi i partecipanti, indicando che il sensore PPG funzionava correttamente. È quindi plausibile che la qualità del contatto degli elettrodi EDA e la lunga durata della sessione abbiano compromesso progressivamente gli altri segnali. Nel complesso, questi risultati giustificano l'utilizzo di dati sintetici in questo lavoro e permettono di individuare con chiarezza i principali aspetti da migliorare nelle sessioni future.

10.3 Contributi metodologici

Rispetto allo stato dell'arte, il lavoro apporta tre contributi metodologici principali.

Architettura ibrida a cascata. Inserire la CPA come stadio di filtraggio a monte del classificatore supervisionato ha permesso di azzerare i falsi positivi nel modello ASD di base. Questo risultato è coerente con il principio enunciato da Apostol et al. [31]: applicata a serie temporali continue, la Change Point Detection riduce sistematicamente i falsi positivi che un classificatore isolato non riesce a evitare.

Baseline dinamica individuale. Calibrare tutte le soglie sulla distribuzione del singolo soggetto, anziché su valori normativi di popolazione, non è una scelta arbitraria: è una necessità dettata dall'ampia variabilità autonoma propria dell'ASD e dalla dispersione inter-individuale documentata nel dataset aeronautico. L'approccio è direttamente estendibile a qualunque contesto in cui la disregolazione autonoma preceda misurabilmente un evento critico.

Framework di trasferibilità cross-dominio. Il risultato forse più rilevante è la dimostrazione che l'architettura CPA+ensemble si trasferisce dalla predizione delle crisi ASD al monitoraggio del MWL aeronautico ricalibrandone i parametri fisiologici e sostituendo il classificatore, senza toccare la logica di base. Questo suggerisce che il framework non sia un modello costruito su misura per questo caso specifico, ma

uno schema metodologico replicabile in qualunque dominio in cui segnali da sensori indossabili anticipino misurabilmente un evento critico.

10.4 Limiti del lavoro

Assenza di un dataset di validazione clinica. Il progetto nasceva con l'obiettivo di raccogliere dati fisiologici reali su soggetti ASD a basso funzionamento durante sessioni di stimolazione sociale e cognitiva, per addestrare e validare il sistema direttamente sul campo. Non è stato possibile costruire un dataset clinico sufficientemente ampio e annotato, e si è dovuto ricorrere a dati sintetici. Questa scelta non riflette una debolezza metodologica, ma le difficoltà concrete ed etiche che la raccolta di dati fisiologici in popolazioni cliniche vulnerabili comporta quasi invariabilmente, come documentato in modo ricorrente nella letteratura di settore. I parametri del modello sintetico, ovvero baseline EDA, forma della rampa pre-crisi, entità del calo HRV, frequenza degli SCR, non sono stati fissati arbitrariamente: sono stati calibrati sulle osservazioni raccolte nella fase sperimentale preliminare, in modo da preservare una corrispondenza qualitativa con la fisiologia reale della popolazione studiata.

Criticità nella qualità del segnale acquisito. L'acquisizione sperimentale condotta con Empatica EmbracePlus (capitolo 9) ha evidenziato che le anomalie nel segnale EDA non sono un artefatto esclusivo del dispositivo E4 usato nella letteratura di riferimento, ma una criticità sistematica dell'acquisizione elettrodermica al polso in condizioni ecologiche. I valori di SCL registrati in entrambi i soggetti sani si sono collocati di oltre un ordine di grandezza al di sotto dei range fisiologici attesi, nonostante l'applicazione del gel conduttivo e la verifica del posizionamento. Questo risultato rafforza la motivazione alla generazione sintetica dei dati di addestramento e indica che, in vista di sessioni su popolazioni cliniche, sarà indispensabile adottare protocolli di verifica della qualità del segnale in tempo reale prima dell'avvio della registrazione.

Popolazione omogenea e non generalizzabilità immediata. I soggetti osservati nella fase di raccolta presentavano un profilo autonomico caratterizzato da iperreattività simpatica agli stimoli sociali e cognitivi, in linea con quanto documentato per l'ASD a basso funzionamento in condizioni di sovraccarico sensoriale. Il modello

è quindi calibrato su questa sottopopolazione specifica: non si trasferisce direttamente a profili autonomici diversi, come soggetti ASD ipo-reattivi o con pattern di risposta atipici, per i quali sarebbe necessaria una raccolta dati dedicata e una ricalibrazione dei parametri.

Dati sintetici con struttura semplificata. Il modello sintetico incorpora variabilità fisiologicamente motivata (rumore tonico, SCR soggetto-specifiche, latenza della risposta HRV e rampe pre-crisi con ampiezze diverse) ma non riproduce il segnale fisiologico reale nella sua interezza. Mancano in particolare gli artefatti da movimento sovrapposti alla componente EDA, la non-stazionarietà a lungo termine tipica delle sessioni cliniche prolungate e la decomposizione tonica-fasica del segnale secondo metodi validati come *cvxEDA*. Le performance riportate vanno quindi lette come stime dell'efficacia del sistema in condizioni controllate: su dati reali è ragionevole aspettarsi una riduzione delle metriche.

Limitazioni della validazione cross-subject. La LOO-CV su 10 soggetti sintetici fornisce una stima non distorta della capacità di generalizzazione inter-soggetto del classificatore all'interno del dominio sintetico, ma non costituisce una validazione su dati reali. In particolare, non permette di quantificare il gap di performance nel passaggio dal dominio sintetico a quello reale, che nella letteratura sul domain adaptation in biosegnali può essere rilevante. Per il modello ASD, la validazione su singolo soggetto sintetico con analisi di robustezza al rumore dimostra la stabilità dell'architettura rispetto a perturbazioni del segnale, ma non è sufficiente a caratterizzare la variabilità inter-soggetto attesa in un contesto clinico reale.

10.5 Prospettive future

I limiti descritti nel paragrafo precedente tracciano in modo abbastanza diretto le direzioni di sviluppo più urgenti. La priorità è la raccolta di un dataset clinico annotato su soggetti ASD a basso funzionamento, con almeno 10–15 partecipanti e sessioni abbastanza lunghe da includere episodi di escalation comportamentale. L'architettura proposta (Change Point Detection non supervisionata seguita da classificazione con matrice di costo asimmetrica) si presterebbe ad essere applicata su questi dati sosti-

tuendo il generatore sintetico con un modulo di acquisizione reale, senza richiedere modifiche strutturali al sistema.

In tale prospettiva, l'esperienza acquisita nel capitolo 9 suggerisce alcuni interventi di protocollo da adottare sistematicamente in sessioni future: verificare la qualità del segnale EDA in tempo reale tramite la CareLab App prima dell'avvio della registrazione, attendere almeno 5 minuti dall'indossaggio per consentire la stabilizzazione termica e del contatto elettrodo, e limitare la durata complessiva delle sessioni per ridurre il rischio di degrado progressivo del segnale. L'adozione di questi accorgimenti è una preconditione necessaria per ottenere dati EDA affidabili su cui addestrare e validare il sistema in contesti reali.

Una seconda direzione riguarda l'estensione a profili autonomici diversi da quello iper-reattivo. Questo richiederebbe dati su sottopopolazioni ASD più eterogenee e l'introduzione di un modulo di identificazione automatica del profilo del soggetto, capace di selezionare dinamicamente le soglie più adeguate prima della fase di classificazione.

La terza direzione è la transizione dall'analisi offline all'elaborazione in streaming. L'implementazione di un buffer circolare FIFO con soglie aggiornate tramite statistiche a finestra scorrevole, descritta nella sezione 7.9, rappresenta il passo tecnico necessario per il deployment su dispositivo embedded. A questo si aggiunge la possibilità di trasferire il modello pre-addestrato su dati sintetici a un nuovo paziente tramite fine-tuning su poche sessioni di calibrazione, approccio validato da Lange et al. [33]: questa strategia è probabilmente la più praticabile per superare la scarsità di dati etichettati nelle popolazioni cliniche rare, consentendo un affinamento incrementale del modello man mano che nuove osservazioni si rendono disponibili.

Bibliografia

- [1] Stephen M. Kanne e Micah O. Mazurek. «Aggression in Children and Adolescents with ASD: Prevalence and Risk Factors». In: *Journal of autism and developmental disorders* 41.7 (2011), pp. 926–937. DOI: 10.1007/s10803-010-1118-4.
- [2] Peizheng Wang, Robert Houghton e Arnab Majumdar. «Detecting and predicting pilot mental workload using heart rate variability: a systematic review». In: *Sensors* 24.12 (2024), p. 3723. DOI: 10.3390/s24123723.
- [3] Matthew S. Goodwin et al. «Predicting aggression to others in youth With autism using a wearable biosensor». In: *Autism research* 12.8 (2019), pp. 1286–1296. DOI: 10.1002/aur.2151.
- [4] Sinéad Lydon, Olive Healy e Martina Dwyer. «An examination of heart rate during challenging behavior in Autism Spectrum Disorder». In: *Journal of developmental and physical disabilities* 25.2 (2013), pp. 149–170. DOI: 10.1007/s10882-012-9324-y.
- [5] Rachel M. Fenning et al. «Sympathetic-parasympathetic interaction and externalizing problems in children with Autism Spectrum Disorder». In: *Autism research* 12.12 (2019), pp. 1805–1816. DOI: 10.1002/aur.2187.
- [6] Karla Conn Welch. «Physiological signals of autistic children can be useful». In: *IEEE Instrumentation & measurement magazine* 15.1 (2012), pp. 28–32.
- [7] June Groden et al. «Assessing cardiovascular responses to stressors in individuals with Autism Spectrum Disorders». In: *Focus on autism and other developmental disabilities* 20.4 (2005), pp. 244–252.
- [8] Matthew S. Goodwin et al. «Cardiovascular arousal in individuals with autism». In: *Focus on autism and other developmental disabilities* 21.2 (2006), pp. 100–123.

- [9] Stephen W. Porges. *The polyvagal theory: Neurophysiological foundations of emotions, attachment, communication, and self-regulation*. New York: W. W. Norton & Company, 2011.
- [10] Shaked Rosenblum, Sharona L. Rab e Roe Admon. «Dynamics in physiological acute stress response trajectories: uncovering latent variability». In: *BMC Psychiatry* 25 (2025), p. 361. DOI: 10.1186/s12888-025-06807-2.
- [11] Stefania Coelli et al. «Modulation of autonomic responses to cognitive tasks under acute mental stress». In: *Scientific Reports* 16 (2026), p. 4878. DOI: 10.1038/s41598-025-34921-4.
- [12] Xiaoru Yu et al. «Exploring physiological stress response evoked by passive translational acceleration in healthy adults: a pilot study utilizing electrodermal activity and heart rate variability measurements». In: *Scientific reports* 14 (2024), p. 11349. DOI: 10.1038/s41598-024-61656-5.
- [13] Azadeh Kushki et al. «A Kalman filtering framework for physiological detection of anxiety-related arousal in children with Autism Spectrum Disorder». In: *IEEE Transactions on biomedical engineering* 62.4 (2015), pp. 990–1000.
- [14] Tales Imbiriba et al. «Wearable biosensing to predict imminent aggressive behavior in psychiatric inpatient youths with autism». In: *JAMA network open* 6.12 (2023), e2348898. DOI: 10.1001/jamanetworkopen.2023.48898.
- [15] David Rosenfield et al. «Change-Point Analysis for longitudinal physiological data: detection of cardio-respiratory changes preceding panic Attacks». In: *Biological psychology* 84.1 (2010), pp. 112–120. DOI: 10.1016/j.biopsycho.2010.01.020.
- [16] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. «Heart rate variability: standards of measurement, physiological interpretation, and clinical use». In: *Circulation* 93.5 (1996), pp. 1043–1065. DOI: 10.1161/01.CIR.93.5.1043.

- [17] Gary G. Berntson et al. «Heart rate variability: origins, methods, and interpretive caveats». In: *Psychophysiology* 34.6 (1997), pp. 623–648. DOI: 10.1111/j.1469-8986.1997.tb02140.x.
- [18] Theodore P. Beauchaine. «Vagal tone, development, and Gray’s motivational theory: toward an integrated model of autonomic nervous system functioning in psychopathology». In: *Development and psychopathology* 13.2 (2001), pp. 183–214. DOI: 10.1017/S0954579401002012.
- [19] S. J. Hutt et al. «Arousal and childhood autism». In: *Nature* 204 (1964), pp. 908–909. DOI: 10.1038/204908a0.
- [20] Austin M. DesLauriers e Caruth F. Carlson. *Your child is asleep: early infantile autism*. Homewood, IL: Dorsey press, 1969.
- [21] Susan R. Leekam et al. «Describing the sensory abnormalities of children and adults with autism». In: *Journal of autism and developmental disorders* 37.5 (2007), pp. 894–910. DOI: 10.1007/s10803-006-0218-7.
- [22] Sarah A. Schoen et al. «Physiological and behavioral differences in sensory processing: a comparison of children with Autism Spectrum Disorder and Sensory Modulation Disorder». In: *Frontiers in integrative neuroscience* 3 (2009), p. 29. DOI: 10.3389/neuro.07.029.2009.
- [23] JP Kootz, B Marinelli e DJ Cohen. «Sensory receptor sensitivity in autistic children: response times to proximal and distal stimulation». In: *Archives of general psychiatry* 38.3 (1981), pp. 271–273. DOI: 10.1001/archpsyc.1981.01780280039003.
- [24] Sakeena Panju et al. «Atypical sympathetic arousal in children with Autism Spectrum Disorder and its association with anxiety symptomatology». In: *Molecular autism* 6 (2015), p. 64. DOI: 10.1186/s13229-015-0057-5.
- [25] Kayisan M. Dalmeida e Giovanni L. Masala. «HRV features as viable physiological markers for stress detection using wearable devices». In: *Sensors* 21.8 (2021), p. 2873. DOI: 10.3390/s21082873.

- [26] Md Abul Hasan et al. «A comprehensive survey and comparative analysis of time series data augmentation in medical wearable computing». In: *PLOS ONE* 20.3 (2025), e0315343. DOI: 10.1371/journal.pone.0315343.
- [27] Ming-Zher Poh, Nicholas C. Swenson e Rosalind W. Picard. «A wearable sensor for unobtrusive, long-term assessment of electrodermal activity». In: *IEEE Transactions on biomedical engineering* 57.5 (2010), pp. 1243–1252. DOI: 10.1109/TBME.2009.2038487.
- [28] Alberto Greco et al. «cvxEDA: A convex optimization approach to electrodermal activity processing». In: *IEEE Transactions on biomedical engineering* 63.4 (2016), pp. 797–804. DOI: 10.1109/TBME.2015.2474131.
- [29] Emily Barbara Prince et al. «The relationship between autism symptoms and arousal level in toddlers with autism spectrum disorder, as measured by electrodermal activity». In: *Autism* 21.4 (2017), pp. 504–508. DOI: 10.1177/1362361316648816.
- [30] Bradley J. Ferguson et al. «Examining the association between electrodermal activity and problem behavior in severe Autism Spectrum Disorder: a feasibility study». In: *Frontiers in psychiatry* 10 (2019), p. 654. DOI: 10.3389/fpsy.2019.00654.
- [31] Elena Simona Apostol et al. «Change point enhanced anomaly detection for IoT time series data». In: *Water* 13.12 (2021), p. 1633. DOI: 10.3390/w13121633.
- [32] P. Zihisire Muke e Adrianna Kozierekiewicz. «Machine learning techniques to improve the cognitive workload classification using multimodal sensors' data». In: *IEEE Access* 13 (2025), pp. 173415–173443. DOI: 10.1109/ACCESS.2025.3616788.
- [33] Lucas Lange, Nils Wenzlitschke e Erhard Rahm. «Generating synthetic health sensor data for privacy-preserving wearable stress detection». In: *Sensors* 24.10 (2024), p. 3052. DOI: 10.3390/s24103052.

- [34] Hassan Ismail Fawaz et al. «Data augmentation using synthetic data for time series classification with deep residual networks». In: *CoRR* abs/1808.02455 (2018). URL: <https://arxiv.org/abs/1808.02455>.
- [35] Farida Sabry et al. «Wearable data generation using time-series generative adversarial networks for hydration monitoring». In: *Proceedings of the 16th International Joint Conference on biomedical engineering systems and technologies (BIOSTEC 2023) – Volume 4: BIOSIGNALS*. SCITEPRESS, 2023, pp. 94–105. DOI: 10.5220/0011757200003414.
- [36] Mathias Benedek e Christian Kaernbach. «A continuous measure of phasic electrodermal activity». In: *Journal of neuroscience methods* 190.1 (2010), pp. 80–91. DOI: 10.1016/j.jneumeth.2010.04.028.
- [37] Ming-Zher Poh et al. «Convulsive seizure detection using a wrist-worn electrodermal activity and accelerometry biosensor». In: *Epilepsia* 53.5 (2012), e93–e97. DOI: 10.1111/j.1528-1167.2012.03444.x.
- [38] Sandra Hödl et al. «Pre-ictal heart rate variability alterations in focal onset seizures and response to vagus nerve stimulation». In: *Seizure: european journal of epilepsy* 86 (2021), pp. 175–180. DOI: 10.1016/j.seizure.2021.02.017.
- [39] Victoria López et al. «Analysis of preprocessing vs. cost-sensitive learning for imbalanced classification. Open problems on intrinsic data characteristics». In: *Expert systems with applications* 39.7 (2012), pp. 6585–6608. DOI: 10.1016/j.eswa.2011.12.043.

Ringraziamenti