POLITECNICO DI MILANO

School of Industrial and Information Engineering
Master of Science in Biomedical Engineering

Deep convolutional networks for concurrent photoplethysmographic and phonocardiographic signal processing

Supervisor: Prof. Pietro Cerveri
Co-Supervisors: Ing. Dario Bovio
Ing. Caterina Salito

Thesis of:
Andrea Dorizza
Student ID: 897261

Academic Year 2018-2019

Sommario

Introduzione

L'ipertensione è anche conosciuta come "l'assassino silenzioso", poiché la maggior parte delle persone che ne sono affette non è a conoscenza del problema in quanto solitamente asintomatico. Nonostante ciò, l'pertensione è uno dei più importanti fattori di rischio per quasi tutte le malattie cardiovascolari e colpisce 1,13 miliardi di persone in tutto il mondo. Viene normalmente tenuta sotto osservazione attraverso metodi non invasivi per il monitoraggio continuo in ambiente domenstico, sfruttando dispositivi oscillometrici che consentono di ottenere numerose misure di pressione in un arco temporale che varia da uno a pochi giorni. Grazie alla misurazione multicampione e alle condizioni quotidiane in cui viene eseguita, questo approccio ha un grande vantaggio: consente la diagnosi di ipertensione da camice bianco e di ipertensione mascherata, fenomeni che consistono rispettivamente in una sopravvalutazione e in una sottovalutazione della pressione quando la misurazione avviene nello studio medico. Tuttavia, questi approcci si basano esclusivamente sulla tecnica oscillometrica tramite sfigmomanometro, che consente solo misurazioni intermittenti, può causare disagio al soggetto esaminato e, sebbene si verifichi mentre gli individui svolgono le loro normali attività quotidiane, richiede che il soggetto sia immobile durante l'inflazione della cuffia, il che è talvolta associato a letture non affidabili. Sebbene esistano altre tecniche non invasive, che prevedano anche un monitoraggio continuo non intermittente, come il metodo di bloccaggio del volume o la tonometria di applanazione dell'arteria radiale, queste sono impossibili da sfruttare nel contesto quotidiano, perché il loro uso è fortemente limitato ad applicazioni specifiche, come i soggetti sottoposti a chirurgia o in unità di terapia intensiva, dal momento che richiedono di avere macchinari costosi e complessi da utilizzare in un ambiente controllato. Da questo contesto emerge la desiderabilità di nuovi sistemi di monitoraggio di pressione continui e non invasivi in grado di garantire la possibilità e i vantaggi delle misurazioni in contesto quotidiano, nonché il comfort dei metodi senza cuffia.

Stato dell'arte

Un passo verso la realizzazione di questo obiettivo è stato fatto nel 2017 da Porro e Vannoni[1, 2] del Politecnico di Milano. Il loro lavoro di tesi, sviluppato in collaborazione con Biocubica S.r.l., ha portato alla realizzazione di un prototipo operativo chiamato Soundi. Il dispositivo, che viene indossato sul petto del soggetto a stretto contatto con la pelle, è in grado di registrare continuamente due segnali cardiovascolari:

il segnale fonocardiografico (PCG), attraverso un sensore di pressione a basso costo, e segnale fotopletismografico (PPG), attraverso un sistema ottico composto da un LED e un fotodiodo. Questi segnali possono essere sfruttati per calcolare il tempo di transito vascolare (VTT) battito per battito, ovvero il ritardo temporale che quantifica il tempo impiegato dall'onda pressoria sanguigna per transitare dal cuore al letto vascolare di un generico sito periferico. Viene calcolato come la differenza tra un punto di riferimento sul PPG, come l'inizio del picco PPG, la pendenza massima positiva o il picco sistolico, e il primo suono cardiaco del PCG (S1), che si verifica all'inizio della sistole ventricolare come effetto della chiusura delle valvole mitrale e tricuspide.[3] Il VTT è stato dimostrato essere correlato alla pressione sanguigna (BP) in numerosi studi[4, 3, 5], permettendo la possibilità di sfruttare questo dispositivo come strumento di monitoraggio pressorio senza cuffia.

Il principio fisiologico su cui si basa questa correlazione può essere modellato assumendo i vasi come i tubi elastici e il flusso sanguigno nelle arterie come una propagazione delle onde pressorie al loro interno. In questo contesto, la velocità dell'onda sfigmica (PWV) è la velocità alla quale l'impulso pressorio si propaga attraverso il sistema circolatorio e viene solitamente utilizzato in ambito clinico per la stima della rigidità e della distensibilità arteriosa, ed è quindi correlata alla BP. La PWV può essere calcolata come il rapporto tra la distanza tra due siti vascolari e il ritardo della propagazione del sangue tra questi due.[6, 7]

$$PWV = \frac{Distance}{Time\,Delay}$$

Questo intervallo temporale può essere stimato in diversi modi, in base al tipo di segnali misurati nei due siti. Se vengono utilizzati solo segnali PPG, questo intervallo di tempo è denominato pulse transit time (PTT) e si riferisce al tempo impiegato da un'onda di pressione per spostarsi tra due siti arteriosi.[6, 7] Per misurarlo, due sistemi ottici vengono posizionati a una distanza fissa, calcolando la distanza temporale di punti corrispondenti nelle onde PPG. Quando ad essere impiegata è una combinazione di ECG e PPG, il ritardo viene denominato pulse arrival time (PAT). Questo si basa sulla distanza temporale tra il picco R dell'ECG e un punto caratteristico del picco del PPG. Il PAT ha però il principale svantaggio di includere sia il PTT che il pre-ejection period (PEP), che è il ritardo tra la depolarizzazione elettrica del ventricolo sinistro e l'inizio della gittata ventricolare.[6, 7]

$$PAT = PPT + PEP$$

Il PEP dipende da una serie di fattori legati al precarico cardiaco, alle frequenza cardiaca e alla contrattilità del cuore. Pertanto, l'inclusione del PEP nel calcolo del tempo di trasmissione dell'onda pulsoria può quindi causare imprecisione.[4] Inoltre, l'impatto del PEP sul PAT complessivo diminuisce con la distanza dal cuore, quindi per misurazioni in sedi prossimali non può essere trascurato.[7] Per limitare l'influenza del PEP, nel 2006 Foo et al.[4] hanno introdotto il vascular transit time, il quale approssima direttamente il PTT essendo indipendente dal PEP. I loro risultati hanno mostrato che le variazioni osservate nel VTT erano correlate sia alla BP sistolica ($R2 = 0{,}820$; $p < 0{,}05$) che alla BP diastolica ($R2 = 0{,}517$; $p < 0{,}05$). Poiché la BP sistolica presentava una correlazione più evidente, questa è stata stimata tramite modelli regressivi

ottenendo risultati positivi nella previsione. Questa relazione tra BP e VTT è stata ulteriormente confermata da numerosi altri studi, che hanno adottato diversi modelli matematici predittivi, sia lineari che non lineari.[3, 5, 8] In generale, per determinare il modello migliore per la stima della BP tramite VTT/PTT/PAT, è essenziale indurre variazioni di BP in un ampio intervallo al fine di avere più campioni per costruire la curva di calibrazione specifica per ciascun soggetto. Variazioni significative in BP possono essere ottenute utilizzando molti metodi[7], come l'attività fisica, ma mentre la sistolica è più propensa a variazioni, quella diastolica è più restia.

Materiali e metodi

L'obiettivo di questa tesi è stato quello di sfruttare i segnali cardiaci di alta qualità registrati da Soundi, quindi PPG e PCG, per calcolare il VTT con un metodo affidabile e computazionalmente poco dispendioso, per poi impiegarlo per stimare la pressione arteriosa basandosi sulle nozioni teoriche e fisiologiche spiegate. Il progetto può essere diviso in tre parti principali, descritte di seguito.

Object detection

La prima parte del progetto è stata mirata a costruire un algoritmo automatico per l'identificazione di tutti i punti di riferimento necessari per il calcolo del VTT. Questo obiettivo è stato raggiunto tramite l'impiego del deep learning, poiché questo consente di elaborare segnali quasi grezzi, riducendo considerevolmente i costi di calcolo richiesti dalla fase di pre-elaborazione, ed è anche potenzialmente meno sensibile al rumore, consentendo di lavorare in modo più efficace con dati reali. Poiché non erano disponibili online set di dati contenenti PPG e PCG simultanei, il primo step del progetto è stata la raccolta di segnali e l'implementazione del dataset da utilizzare per le fasi di training e test della rete neurale. Il protocollo per l'acquisizione dei dati è stato suddiviso in una fase di setup del dispositivo, in cui Soundi è stato posizionato sul torace del soggetto ed è stata verificata la qualità del segnale, e in una fase di acquisizione dei dati eseguita con due diverse modalità, al fine di aumentare la variabilità della frequenza cardiaca nel dataset. La prima modalità implicava la registrazione per 15 minuti con il soggetto a riposo, mentre la seconda prevedeva, sì la registrazione per 15 minuti, ma dopo lo svolgimento di un'intensa attività fisica da parte del soggetto, consistente nel correre su un tapis roulant a 10 km/h per 10 minuti. Il database finale era costituito da circa 5 ore e 10 minuti di segnali e coinvolgeva un totale di 15 volontari adulti, di cui 6 femmine e 9 maschi. L'età dei partecipanti era compresa tra 24 e 53 anni e nessuno presentava anomalie cardiovascolari note. Poiché l'apprendimento supervisionato necessita di segnali etichettati, è stato implementato un algoritmo di marcatura automatica per rilevare i punti di riferimento: il piede è stato identificato come il minimo locale del PPG, la pendenza massima come il massimo locale della derivata prima del PPG dopo ciascun piede, il picco sistolico come il massimo locale del PPG, il picco S1 come il massimo locale prima di ogni piede del PPG. Dopo l'etichettatura, sono stati utilizzati segnali quasi grezzi per creare i dataset di test e training al fine di migliorare la resistenza al rumore della rete. I campioni di input sono stati realizzati suddividendo ogni registrazione in segmenti più piccoli, ciascuno lungo 600 campioni. Poiché la

frequenza di campionamento era di 400 Hz, ogni segmento corrispondeva a 1,5 secondi di registrazione. La scelta della lunghezza è stata opportunamente oculata in quanto compromesso che tiene conto di diversi fattori, come il numero di segmenti necessari per addestrare con successo la rete, l'effetto sulla dimensione della rete e sui bisogni computazionali, la lunghezza minima per includere almeno un ciclo cardiaco per segmento. La struttura dell'output è stata concepita come una riorganizzazione dell'etichettatura ed era strettamente correlata all'architettura della rete e al suo funzionamento. Ogni segmento lungo 600 campioni è stato virtualmente diviso in cinque finestre, ciascuna di 120 campioni, corrispondenti a 0,3 secondi. In ognuna di queste finestre avrebbe potuto essere presente al massimo un punto di riferimento per tipo, con una posizione specifica all'interno della finestra. Per ciascuna finestra è stato definito un bit di classificazione binaria che indicasse la presenza di un'etichetta nella finestra e un valore che indicasse la sua posizione normalizzata rispetto alla lunghezza della finestra stessa. Il numero di finestre è stato scelto in modo tale da avere il minor numero di output possibili e per evitare di avere più di un'etichetta per tipo nella stessa finestra. Il set di dati è stato quindi diviso in un set di addestramento, un set di ottimizzazione e un set di test, rispettivamente di 20.000 segmenti, 1.000 segmenti, 1.000 segmenti. Il set di training è stato poi raddoppiato da un processo di data augmentation, consistente nella creazione di copie rumorose dei campioni originali per migliorare la capacità della rete di ignorare il rumore e ridurre l'overfitting. Sia per il PPG che per il PCG sono state testate diverse architetture di rete, variando elementi quali numero di strati, numero di filtri, velocità di apprendimento, dimensione dei filtri, batch size. Inoltre, è stato testato anche il numero di tipi di riferimenti del PPG che la rete era in grado di rilevare contemporaneamente, nonché la possibilità di far funzionare le reti in sequenza, alimentando la rete del PCG anche con l'output simultaneo della rete del PPG. Il processo di addestramento è stato eseguito dopo aver definito una funzione di ottimizzazione appositamente progettata:

$$Custom\ Loss = Identification\ Loss + Localization\ Loss$$

$$Identification\ Loss = \frac{1}{N}\sum_{i=1}^{N} |y_{i,true}^{id} - y_{i,pred}^{id}|$$

$$Localization\ Loss = \frac{1}{N_{TP}}\sum_{i=1}^{N} \left(T_{id}(y_{i,pred}^{id}) \times y_{i,true}^{id} \times |y_{i,true}^{loc} - y_{i,pred}^{loc}|\right)$$

$$T_{id}(a) = \begin{cases} 1, & \text{if } a \geq 0.5 \\ 0, & \text{if } a < 0.5 \end{cases}$$

dove la identification loss è calcolata come la distanza media assoluta tra la confidenza della rete nella presenza di un punto di riferimento e il valore binario corrispondente. In questo caso, la previsione fatta dalla rete misura quanto questa sia sicura della presenza di un punto di riferimento specifico. La localization loss invece misura l'errore di previsione della posizione del punto, quindi viene calcolata solo quando il punto predetto è effettivamente presente in una finestra specifica. Per questo motivo è stata calcolata come l'errore medio assoluto di localizzazione quando era effettivamente presente il punto di riferimento predetto, ma ridotto a zero quando assente. N_{TP} é il numero totale di reali positivi.

Estrazione delle feature

La seconda parte è stata dedicata alla estrazione di feature rilevanti, sfruttando i punti identificati con la rete neurale. Poiché il miglior punto di riferimento del PPG per il calcolo del VTT e la conseguente stima della BP non è stato ancora studiato diffusamente, il VTT è stato calcolato in tre modi diversi, in base al punto di riferimento del PPG impiegato, al fine di indagare su queste relazioni. Per ogni campione, il VTT_i è stato calcolato come la differenza tra t_i, dove i è un tipo di riferimento, e il $t_{S1} < t_i$ più vicino. Anche la frequenza cardiacaè stata estratta in modo simile: per ogni tipo di riferimento e per ogni campione, HR_i è stata calcolata come la differenza tra t_j e il $\widetilde{t_j} < t_j$ più vicino.

Stima della pressione

La terza parte riporta un caso d'uso degli algoritmi implementati. In questa fase sono stati sviluppati alcuni algoritmi di stima della pressione arteriosa sistolica per mappare le caratteristiche estratte ai corrispondenti valori di pressione sanguigna, sfruttando diversi modelli teorici. Come per il primo database, anche per quello per la regressione il protocollo di acquisizione può essere suddiviso in due fasi principali: la fase di configurazione del sistema e la fase di acquisizione dei dati. La fase di setup è stata sostanzialmente identica, tranne per il fatto che il soggetto ha dovuto indossare l'oscillometro GIMA ABPM sul braccio sinistro. La fase di acquisizione è stata invece profondamente diversa. Poiché per addestrare un modello di regressione che prevede la BP a partire dal VTT sono necessarie diverse coppie di valori corrispondenti, la variabilità intra-soggetto è stata garantita inducendo una variazione sensibile della BP attraverso un esercizio fisico supervisionato. Il soggetto ha dovuto sedersi su un cicloergometro per 60 minuti, indossando sia Soundi che l'holter pressorio. Gli è stato chiesto di pedalare per 3 minuti a velocità costante di 15 km/h, quindi di fermarsi per 2 minuti. All'inizio della pausa, è stata eseguita una misurazione di pressione, della durata di circa 30 secondi. Il processo descritto è stato ripetuto per tre volte di 10 minuti ciascuna, aumentando l'intensità dell'esercizio ogni volta. Il dataset è stato creato isolando i segmenti di segnale dopo l'inizio di ogni pausa, in una finestra temporale di 30 secondi. All'esperimento ha partecipato un soggetto maschio di 25 anni, perciò il dataset era composto da 12 segmenti di 30 secondi ciascuno per ciscun tipo di segnale e 12 misurazioni numeriche simultanee. La rete è stata utilizzata per estrarre sia il VTT che l'HR per ciascun segmento, poi è stata calcolata la media dei valori di tutte feature estratte appartenenti allo stesso segmento. Pertanto, il risultato finale era una sequenza soggetto-specifica di 12 set di feature, ciascuno contenente tre valori di VTT e un valore di HR, associato a un numero uguale di set di pressione sanguigna corrispondenti, ciascuno con un valore di pressione sistolica e uno di diastolica. Queste tuple sono quindi state utilizzate per diversi modelli di regressione lineare, sia semplice che multipla, che descrivevano una relazione lineare, reciproca o logaritmica tra feature e BP. Data la dimensione limitata del dataset, per sfruttarlo al meglio i modelli sono stati valutati adottando un approccio di cross-validazione leave-one-out.

Risultati

L'architettura di rete con le migliori prestazioni è essenzialmente la stessa sia per il PPG che per il PCG, eccetto per il numero di canali di input e di output. $C_{\text{input, PPG}}$ è uguale a 1, poiché riceve solo i segmenti di PPG. $C_{\text{output, PPG}}$ è uguale a 3, poiché effettua contemporaneamente più identificazioni. $C_{\text{input, PCG}}$ è uguale a 4, poiché riceve in ingresso i segmenti di PCG e l'output dalla rete PPG. Infine, $C_{\text{input, PCG}}$ è uguale a 1, poiché rileva solo i suoni S1.

Le tabelle 1 e 2 mostrano i risultati della rete secondo le metriche standard per la classificazione binaria, mentre nella tabella 3 sono riportati l'errore assoluto e l'errore quadratico medio per l'attività di localizzazione, espressi in millisecondi rispetto al riferimento calcolato algoritmicamente.

Label type	**Accuracy**	**Bal. Acc.**	**Specificity**
PPG_{Foot}	0.97	0.97	0.98
$\text{PPG}_{\text{Max Slope}}$	0.94	0.94	0.95
PPG_{Peak}	0.85	0.85	0.89
PPG_{All}	0.92	0.92	0.94
PCG	0.97	0.97	0.97

Tabella 1: *Metriche standard per problemi di classificazione binaria impiegate per valutare i rusultati della identificazione sul test set.*

Label type	**Precision**	**Recall**	**F1 score**
PPG_{Foot}	0.95	0.95	0.95
$\text{PPG}_{\text{Max Slope}}$	0.90	0.91	0.91
PPG_{Peak}	0.75	0.75	0.75
PPG_{All}	0.86	0.87	0.87
PCG	0.94	0.96	0.95

Tabella 2: *Metriche standard per problemi di classificazione binaria impiegate per valutare i rusultati della identificazione sul test set.*

Label type	**MAE [ms]**	**RMSE [ms]**
PPG_{Foot}	2.83	12.30
$\text{PPG}_{\text{Max Slope}}$	4.49	15.59
PPG_{Peak}	8.96	26.40
PPG_{All}	5.43	19.08
PCG	3.66	17.08

Tabella 3: *Metriche impiegate per valutare i rusultati della localizzazione sul test set. L'errore è associato alla sua deviazione standard.*

Di seguito sono riportati i modelli di regressione con le prestazioni migliori tra quelli testati e nella tabella 4.12 la loro performance. PC_{i} rappresenta la i-esima componente principale ottenuta tramite anlisi delle componenti principali.

- Model A: $SBP = a \times \left(\frac{1}{3}\sum_{i\in(foot,maxslope,peak)} VTT_i\right)^{-1} + b$
- Model B: $SBP = a \times 1/VTT_{maxslope} + b$
- Model C: $SBP = a \times \log(VTT_{maxslope}) + b$
- Model D: $SBP = \sum_{i=1}^{N_{PC}} a_i \times PC_i(VTT_{foot}^{-1}, VTT_{maxslope}^{-1}, VTT_{peak}^{-1}) + b$

Model	**MAE [mmHg]**	**MSE [mmHg^2]**	**R^2**
A	1.81	5.30	0.82
B	1.81	5.37	0.81
C	1.91	5.60	0.81
D	1.97	5.74	0.81

Tabella 4: *Modelli di regressione lineare semplice e multipla e la loro performance sul dataset paziente-specifico.*

Discussione

Il problema è stato affrontato come un problema di object detection traendo ispirazione da alcuni degli algoritmi di deep learning più avanzati che rappresentano lo stato dell'arte, come YOLO [9]. Il rilevamento di oggetti implica due sottoattività: la classificazione di un evento e localizzazione dello stesso all'interno del segnale. Nonostante queste attività vengono di solito svolte separatamente, classificando prima l'oggetto e poi identificandone la posizione attraverso una laboriosa post-elaborazione, in questo progetto di tesi le componenti sono state accorpate in un'unica, potente, flessibile rete neurale convoluzionale ottimizzata per l'elaborazione di segnali monodimensionali. Il problema di object detection è stato quindi riformulato come un singolo problema ibrido di classificazione-regressione, riducendo i costi di calcolo e il tempo necessario per ciascuna predizione.

L'architettura finale può essere divisa concettualmente in due parti con scopi diversi. La prima parte, composta dalla sequenza di layer convoluzionali intervallati da layer di maxpooling, funge da estrattore di feature, identificando pattern caratteristici nei dati. Questa parte è caratterizzata da elementi peculiari, quali la presenza di layer convoluzionali monodimensionali per ridurre lo spazio delle feature rispetto ai layer precedenti, combinando le feature attraverso tutte le feature maps. La seconda parte, costituita da un singolo dense layer finale che fornisce l'output, è responsabile della localizzazione degli oggetti identificati. Il formato dell'output è progettato in modo tale che il segmento del segnale di ingresso sia virtualmente diviso in cinque finestre equidimensionali. Se un oggetto cade in una finestra, quella finestra è responsabile per il rilevamento di quell'oggetto. Ogni finestra è associata a due neuroni di output per ciascun tipo di etichetta: uno che esprime la posizione normalizzata dell'etichetta all'interno del segmento, l'altro un punteggio di confidenza per questa identificazione, che comunica quanto il modello sia certo che il segmento contenga un oggetto. Lo strato

finale può avere una dimensione variabile in base al numero di oggetti da identificare, ognuno dei quali costituisce un diverso canale di output C_{output}. Il suo scopo è di garantire la massima flessibilità riguardo al numero di etichette rilevabili simultaneamente. Il fatto che le prestazioni della rete su ciascuna etichetta fossero comparabili quando testate sul rilevamento di più oggetti rispetto al caso di un singolo oggetto significa che le feature estratte dalla prima parte della rete sono abbastanza generali da descrivere interamente le caratteristiche morfologiche del segmento di segnale in input. Inoltre, è stata dimostrata la possibilità di alimentare la rete con un input multicanale, in cui ciascun canale C_{input} rappresenta un segmento sincrono di segnale. È stata anche dimostrata l'efficacia di questo approccio: quando la rete per il riconoscimento di S1 è stata alimentata con etichette rilevate sul PPG nella configurazione in sequenza, l'accuratezza della predizione e le sue capacità di localizzazione sono migliorate in modo significativo. Questo espediente consente di sfruttare efficacemente ed immediatamente le informazioni provenienti dalla rete del PPG al fine di migliorare la capacità della rete di discriminare tra i suoni S1 e S2 che possono essere molto simili, specialmente nei segnali rumorosi.

Per quanto riguarda la valutazione sul set di test, si può notare che sono stati ottenuti risultati eccellenti soprattutto per quanto riguarda la detection del piede e di S1. Infatti, nella attività di identificazione, sia la rete PPG che quella PCG hanno raggiunto il 97 % di accuratezza bilanciata sulle etichette menzionate, mentre sul rilevamento della pendenza massima questa era del 94 %. Una prestazione peggiore è stata raggiunta sul rilevamento del picco di PPG, poiché la accuratezza bilanciata era dell'85%. Un'analisi statistica più approfondita ha rivelato che il numero di falsi negativi era uguale al numero di falsi positivi, quindi la rete produce un numero assoluto corretto di identificazioni, ma talvolta in una finestra sbagliata. Il motivo di questa erosione di capacità deve essere ricercato nella qualità dell'etichettatura. Infatti, mentre nei segnali di alta qualità il picco sistolico è facilmente riconoscibile, sui dati reali questo può fondersi con quello diastolico. Pertanto, l'algoritmo di etichettatura automatica sviluppato in questo progetto, sebbene abbastanza robusto, non è stato sempre in grado di localizzare correttamente i picchi, causando rumore eccessivo nel training set che ha portato la rete a localizzare erroneamente le etichette. Ad ogni modo, la solidità dell'algoritmo di deep learning è evidente nei casi in cui la rete è stata addestrata con campioni consistenti.

Un'altra caratteristica della rete che merita di essere approfondita è il modo in cui i segmenti vengono elaborati per eseguire un rilevamento. Il fatto che la rete elabori l'intero segmento per eseguire il rilevamento significa che utilizza le caratteristiche di tutto il campione come contesto per predire la posizione delle etichette, ragionando a livello globale sull'intero segmento e considerando tutti gli oggetti presenti, tenendo conto anche della presenza codificata nei dati di etichette contigue.

Un altro punto che richiede una menzione è il livello di pre-elaborazione dei segnali di input con cui la rete è in grado di lavorare. In questo progetto è stata dimostrata la possibilità di utilizzare anche segnali quasi grezzi, riducendo i costi di calcolo e le risorse temporali necessarie per eseguire ciascuna detezione. La presenza di rumore nei dati di training aumenta anche la robustezza della rete, fungendo da fattore di regolarizzazione, come dimostrato da un test eseguito di sensibilità al rumore.

Per quanto riguarda i modelli di stima della pressione arteriosa sistolica, l'etichetta

Label	Study	MAE [mmHg]	R^2
Foot	This work	2.47±2.41	0.75
	[5]	1.14±14.94	0.71
Max slope	This work	1.81±1.36	0.81
	[5]	7.37±11.09	0.84
Peak	This work	2.40±1.15	0.75
	[5]	7.47±11.08	0.84

Tabella 5: *Confronto tra i risultati ottenuti sfruttando il modello A per la pressione sistolica in questo progetto e quelli ottenuti da Esmaili et al. in [5].*

di pendenza massima e l'etichetta del piede sono state prevedibilmente le più utili per la stima della pressione arteriosa. Poiché meno affette dall'errore di localizzazione, sono state in grado di fornire feature più affidabili rispetto al picco. I quattro modelli più promettenti tra quelli testati hanno mostrato un'evidente correlazione lineare tra valori stimati e target, con un coefficiente di determinazione maggiore di 0,80, mentre l'errore medio assoluto e l'errore quadratico medio erano inferiori a 2 mmHg e 6 $mmHg^2$, rispettivamente.

Il modello B, validato in letteratura da Esmaili et al. in [3], propone una relazione reciproca tra SBP e VTT e in questo progetto trova conferma del suo valore, in particolare per la valutazione della pressione arteriosa sistolica. Nella tabella 5 sono riportati i risultati dei due studi per questo modello, tenendo in considerazione anche i diversi punti di riferimento utilizzati. Vale la pena notare che il coefficiente di determinazione è simile per il piede e la pendenza massima, molto diverso per il picco. Risultati molto simili sono stati ottenuti in questo progetto anche con il modello C, che descrive una relazione logaritmica tra VTT e SBP.

Risultati di previsione addirittura superiori sono stati ottenuti sfruttando il modello A, ottenendo un coefficiente di determinazione pari a 0,82 e di conseguenza un errore di stima leggermente inferiore. Questo modello, come modello B, descrive una relazione reciproca tra VTT e SBP, ma qui l'intervallo temporale viene misurato come una media di tutti i VTT calcolati, vale a dire VTT_{Foot}, $VTT_{Max\ slope}$ e VTT_{Peak}. Si può ipotizzare che il calcolo del valore medio riduca parzialmente il rumore intrinseco ai VTT elementari, riducendo la varianza della feature estratta. Questo modello sfrutta al massimo le funzionalità della rete, poiché impiega contemporaneamente ogni oggetto rilevato dall'algoritmo.

Conclusioni

In questo lavoro è stata progettata e implementata una rete neurale rapida, flessibile e precisa per il rilevamento di oggetti in serie temporali. La stessa rete è in grado di riconoscere e localizzare più eventi contemporaneamente con quasi lo stesso costo computazionale. Questa capacità di svolgere compiti diversi contemporaneamente la rende valida ed efficiente per scopi diversi nei campi e contesti più disparati, ad esempio per rilevare complessi QRS nei segnali ECG. Inoltre, la possibilità di utilizzare la stessa architettura di rete con un diverso numero di canali di ingresso apre un'ulteriore varietà di possibilità e casi d'uso. Ad esempio, la rete può essere utilizzata per

elaborare contenuti sincroni di frequenza e tempo. Nonostante il rilevamento di punti sia un compito semplice, questo può costiruire la chiave per abilitare diversi compiti più complessi, come la segmentazione dei segnali. In questa tesi è stata fornita anche una dimostrazione di casi d'uso e delle potenzialità della rete. Le feature estratte utilizzando la rete sono state impiegate con successo per stimare la pressione sistolica, sfruttando e rafforzando i semplici modelli di regressione impiegati in altri paper in contesti simili. Sebbene siano necessari ulteriori studi per indagare su questi dati, i risultati sono molto incoraggianti e coerenti con le aspettative, nonché con i risultati presenti in letteratura.

Abstract

Introduction

Hypertension is also known as "the silent killer", since most affected people are unaware of the problem because it may be totally asymptomatic. Despite its quiet and silent behavior, high blood pressure is one of the strongest risk factors for almost all different cardiovascular diseases and affects 1.13 billion people worldwide. Usually, it is monitored through non-invasive, out-of-office, BP measurement consisting in using either home blood pressure monitoring (HBPM) or ambulatory blood pressure monitoring (ABPM), which exploit oscillometric devices to obtain several BP measures during a period of few days. Because of the multi-sampling measurement approach and the daily life conditions in which it is performed, a major advantage regarding both ABPM and HBPM is that they enable the diagnosis of white-coat and masked hypertension, phenomenons consisting of a BP overestimation and underestimation, respectively, when measurement occurs in doctor's office. Moreover, ABPM offers a series of advantages with respect to HBPM, such as a stronger prognostic evidence, night-time readings, short-term BP variability information. Anyway, these approaches always rely on cuff-based oscillometric technique, that is intermittent, can cause discomfort in the examined patient and, although ABPM occurs while individuals go about their normal daily activities, the subject is required to keep his arm still while the cuff is inflating to avoid excessive motion, which is associated with unobtainable or artefactual readings. Although other non-invasive, even continuous techniques exist, like volume clamp method or radial artery applanation tonometry, these are impossible to exploit for ABPM because their use is strongly limited to specific applications, such as subjects undergoing surgery or in intensive care units, since they require to have expensive and complex machinery to be used in a controlled environment. From this context emerges the desirability of new continuous, non-invasive, BP monitoring systems able to guarantee the possibility and the advantages of ABPM, as well as the comfort of cuffless methods.

State of the art

A step towards realizing this goal was taken in 2017 by Porro and Vannoni[1, 2] from Politecnico di Milano. Their thesis work, raised as a collaboration with Biocubica S.r.l., led to an operational prototype named Soundi. The device, that should be worn on the subject's chest in close contact with his skin, is able to continuously record two cardiovascular signals: phonocardiographic signal (PCG), through a low-cost pressure

sensor, and photoplethysmographic signal (PPG), through an optical system composed by a LED and a receiving photodiode. These signals can be exploited to calculate the beat-to-beat vascular transit time (VTT), a time delay measuring the time spent by the blood pressure wave to transit from the heart to the vascular bed of a generic peripheral site. It is computed as the difference between a reference point on the PPG, such as the onset of the PPG peak, the maximum positive slope or the systolic peak, and the first cardiac sound of the PCG (S1), occurring at the onset of ventricular systole as effect of the closure of the mitral and tricuspid valves.[3] VTT has been proved correlated to blood pressure (BP) in several studies[4, 3, 5], potentially enabling the possibility to exploit this device as a cuff-less ABPM instrument.

The physiological principle on which this correlation is based can be modeled by assuming vessels like elastic tubes and the blood flow in the arteries as the propagation of pressure waves inside them. In this framework, pulse wave velocity (PWV) is the velocity at which the blood pressure pulse propagates through the circulatory system and it is used clinically as a measure of arterial stiffness and arterial distensibility, thus is related to BP. It can be computed as the ratio of the distance between two vascular sites and a time delay in blood propagation.[6, 7]

$$PWV = \frac{Distance}{Time\,Delay}$$

This time delay can be estimated in several ways, according to the type of signals measured at the two sites. If only PPG signals are used, this time interval is named pulse transit time (PTT) and it refers to the time taken by a pressure wave to travel between two arterial sites.[6, 7] To measure it, two optical systems are placed at a fixed distance, calculating the temporal shift of corresponding points in PPG waves. When a combination of ECG and PPG is used, the time delay is named pulse arrival time (PAT). It is based on the time difference between the R-peak of ECG and a characteristic point of PPG peak. PAT has the main disadvantage of including both the PTT and the pre-ejection period (PEP), which is the time delay between the electrical depolarization of the left ventricle and the start of the mechanical ventricular ejection.[6, 7]

$$PAT = PPT + PEP$$

PEP depends upon a number of factors related to the preload, HR and contractility of the heart. Therefore, the inclusion of PEP in the time computation of pulse wave transmission along the vessel can then induce inaccuracy due to these factors.[4] Furthermore, the impact of PEP on the overall PAT decreases with distance from the heart, thus for near measurement locations it can not be neglected and should be taken into account.[7] To help limit the influence of PEP, in 2006 Foo et al.[4] introduced the vascular transit time, which approximates the PTT being independent of PEP. Their results showed that observed VTT changes were related to both systolic BP ($R2 = 0.820$; $p < 0.05$) and diastolic BP ($R2 = 0.517$; $p < 0.05$). As systolic BP had the strongest correlation, a regression equation was formulated to associate the two parameters obtaining positive results in prediction. This possibility has been further confirmed by several other studies, by adopting different mathematical models for prediction, both linear and non-linear.[3, 5, 8] In general, to find the best fitting curve for VTT/PTT/PAT-based BP estimation, it is essential to vary BP over a wide range in

order to have more points to construct the subject-specific calibration curve. Significant variations in SBP can be achieved by using many methods[7], such as physical activity, but while SBP is more reactive, DBP is less likely to change.

Materials and methods

The objective of this thesis was to exploit the high-quality cardiac signals recorded by Soundi, namely PPG and PCG, to compute vascular transit time in a reliable, computationally-undemanding, robust way with the aim of using it to estimate arterial blood pressure by leveraging the theoretical, physiological basis shown. In order to accomplish the task, the project was divide in three main parts, described in the following.

Object detection

The first part aims at building an automatic algorithm for the identification of all the reference points needed for VTT computation. To accomplish this task a deep learning approach was chosen, since it allows to process almost-raw signals, considerably reducing the computational costs required by the pre-processing phase, and is also potentially more robust to noise, allowing to work more effectively with real-world data. Since no datasets containing concurrent PPG and PCG were available online, the first step of the project was the collection of signals and the implementation of the dataset to be used for network training and testing phases. The protocol for the data acquisition was divided in system setup phase, where Soundi was positioned on the subject's chest and signal quality was checked, and data acquisition phase that was executed with two different modalities, in order to increase the HR variability in the dataset. The first modality implied recording for 15 minutes with the subject at rest, while the second one contemplate recording for 15 minutes as well, but after the subject performed an intense physical activity, consisting in running on a treadmill at 10 km/h for 10 minutes. The final database was constituted by about 5 hours and 10 minutes of signals and involved a total of 15 adult volunteers, of which 6 females and 9 males. The age of participants ranged 24-53 years old and no one presented known cardiovascular anomalies. Since supervised learning required labeled signals, an automatic labeling algorithm to detect reference points was implemented: foot was determined as the local minimum of the PPG, maximum slope as the local maximum after each foot of the first derivative PPG, systolic peak as the local maximum of PPG, PCG S1 peak as the local maximum before each PPG foot. After the labeling, almost-raw signals were used to create the network training and testing sets to improve CNN robustness to noise. The inputs were realized by splitting each signal recording in smaller chunks, each 600 samples long. Since sampling frequency was 400 Hz, each chunk corresponded to 1.5 seconds of recording. The choice of the length was opportunely calibrated, as a compromise accounting for several factors, such as the number of chunks needed to train successfully the network, the effect on the network dimension and on computational necessities, the minimum length to include at least on cardiac cycle per chunk. Output target structure was built by a rearrangement of the labeling and was strictly related to the network architecture and its functioning. Each 600 long sample was virtually

divided in five windows, 120 each, thus 0.3 seconds. In each of these windows, at most one reference point per type could have been present, with a specific location inside the window. Thus for each window was defined a binary classification bit indicating the presence of a label in the window and a value indicating its normalized position with respect to the length of the window itself. The number of windows was chosen to have as few network outputs as possible and to avoid having more than one label per type in the same window. The dataset were then divided in a training set, a tuning set and a test set, of 20,000 samples, 1,000 samples, 1,000 samples, respectively. Training set was then doubled by a data augmentation process, consisting in creating noisy copies of the original samples to improve network ability to ignore noise in data and reduce overfitting. For both PPG and PCG several network architectures were tested, varying elements such as number of layers, number of filters, learning rate, filter dimension, training batch size. Furthermore, the number of PPG label types the network could detect at the same time was also tested, as well as the possibility to make the networks working in chain, by feeding the PCG network also with concurrent PPG network outputs. The training process was performed after defining a custom loss functions:

$$Custom\ Loss = Identification\ Loss + Localization\ Loss$$

$$Identification\ Loss = \frac{1}{N}\sum_{i=1}^{N} |y_{i,true}^{id} - y_{i,pred}^{id}|$$

$$Localization\ Loss = \frac{1}{N_{TP}}\sum_{i=1}^{N} \left(T_{id}(y_{i,pred}^{id}) \times y_{i,true}^{id} \times |y_{i,true}^{loc} - y_{i,pred}^{loc}|\right)$$

$$T_{id}(a) = \begin{cases} 1, & \text{if } a \geq 0.5 \\ 0, & \text{if } a < 0.5 \end{cases}$$

Where the identification loss is the mean absolute distance between the prediction of the presence of a reference point and the corresponding true value, ergo the mean absolute error (MAE). In this case, the prediction made by the NN measures how confident it is of the presence of a specific landmark. The localization loss instead measures the prediction error of the point location, thus should be computed only when the predicted point is actually present in a specific window. For this reason it was defined as the MAE of localization when a predicted reference point was present, but shrunk towards zero when absent. N_{TP} is the total number of true positives.

Feature extraction

The second part was dedicated to the crafting of relevant features by exploiting the points detected with the neural network. Since the best PPG reference point for VTT computation and BP estimation has not been studied yet, VTT was computed in three different ways, according to the PPG reference point used, to investigate these relationships. For each sample, VTT_{i} was computed as the difference between t_{i}, where i is a PPG reference type, and the nearest $t_{\mathrm{S1}} < t_{\mathrm{i}}$. Also the heart rate was similarly extracted: per each reference type of each sample, HR_{i} was computed as the difference between t_{j} and the nearest $\widetilde{t_{\mathrm{j}}} < t_{\mathrm{j}}$.

BP estimation

The third part presents a use case of the algorithms implemented. In this phase some systolic blood pressure estimation algorithms were implemented to map the extracted features to the corresponding blood pressure values, leveraging several theoretical model.

As for the NN database, also for regression database the acquisition protocol can be divided in two main phases: again, the system setup phase and data acquisition phase. The system setup phase was basically identical, except for the fact that the subject was required to wear the GIMA ABPM oscillometer on the left arm. The acquisition phase was instead profoundly different. Since to train a regression model predicting BP from VTT several couples of corresponding values are needed, intra-subject variability was guaranteed by inducing appreciable BP variation through a supervised physical exercise. The subject was asked to sit on a cycloergometer for 60 minutes, wearing both Soundi and the holter. He was asked to cycle for 3 minutes at constant speed of 15 km/h, then to stop for 2 minutes. At the beginning of the pause, a BP sampling, lasting about 30 seconds, was executed. The described process was repeated for three times of 10 minutes each, increasing the intensity of the exercise at each time. The dataset was built by extracting the signal segments after the beginning of each pause, in a temporal window of 30 seconds. One male subject of 25 years old participated at the experiment, thus the dataset was composed by 12 segments of 30 seconds each per signal type and 12 simultaneous numeric measurements. The network was used to extract both VTTs and HRs per each segment, then the average of features values belonging to the same segment was computed. Thus, the final result was a subject-specific sequences of 12 feature sets, each including three VTT computations and one HR value, associated to an equal number of corresponding blood pressure sets, each including systolic and diastolic blood pressure values. This tuples were then used to several both simple and multiple linear regression models, describing a linear, reciprocal or logarithmic relationship between features and BP. Since the dimension of the dataset was limited, to make the most of it models were evaluated by adopting a leave-one-out cross-validation approach.

Results

The best performing network architecture was the same for PPG and PCG, but the number of input and output channels was different. $C_{\text{input, PPG}}$ was equal to 1, receiving only PPG signal segments. $C_{\text{output, PPG}}$ was equal to 3, since simultaneously output multiple labels, namely foot, maximum slope and systolic peak. $C_{\text{input, PCG}}$ was equal to 4, since was receiving the PCG segments and the output labeling from the PPG network. Finally, $C_{\text{input, PCG}}$ was equal to 1, since it was detecting only the S1 sounds.

Table 6 and 7 shows the standard metrics for binary classification tasks, while in Table 8 is reported the absolute and the root mean squared error for the localization task, expressed in milliseconds with respect to the target, the output of automatic detection algorithm.

Label type	Accuracy	Bal. Acc.	Specificity
$\mathrm{PPG_{Foot}}$	0.97	0.97	0.98
$\mathrm{PPG_{Max\ Slope}}$	0.94	0.94	0.95
$\mathrm{PPG_{Peak}}$	0.85	0.85	0.89
$\mathrm{PPG_{All}}$	0.92	0.92	0.94
PCG	0.97	0.97	0.97

Tabella 6: *Standard metrics for binary classification evaluating the identification results on the test set.*

Label type	Precision	Recall	F1 score
$\mathrm{PPG_{Foot}}$	0.95	0.95	0.95
$\mathrm{PPG_{Max\ Slope}}$	0.90	0.91	0.91
$\mathrm{PPG_{Peak}}$	0.75	0.75	0.75
$\mathrm{PPG_{All}}$	0.86	0.87	0.87
PCG	0.94	0.96	0.95

Tabella 7: *Standard metrics for binary classification evaluating the identification results on the test set.*

Label type	MAE [ms]	RMSE [ms]
$\mathrm{PPG_{Foot}}$	2.83	12.30
$\mathrm{PPG_{Max\ Slope}}$	4.49	15.59
$\mathrm{PPG_{Peak}}$	8.96	26.40
$\mathrm{PPG_{All}}$	5.43	19.08
PCG	3.66	17.08

Tabella 8: *Metrics evaluating the localization results on the test set. Error is reported with its standard deviation.*

In the following are reported the best performing regression models among the tested ones, in Table 4.12 their performances. $\mathrm{PC_i}$ represents the i^{th} component obtained through principal component analysis.

- Model A: $SBP = a \times \left(\frac{1}{3}\sum_{i\in(foot,maxslope,peak)} VTT_i\right)^{-1} + b$
- Model B: $SBP = a \times 1/VTT_{maxslope} + b$
- Model C: $SBP = a \times \log(VTT_{maxslope}) + b$
- Model D: $SBP = \sum_{i=1}^{N_{PC}} a_i \times PC_i(VTT_{foot}^{-1}, VTT_{maxslope}^{-1}, VTT_{peak}^{-1}) + b$

Model	MAE [mmHg]	MSE [mmHg^2]	R^2
A	1.81	5.30	0.82
B	1.81	5.37	0.81
C	1.91	5.60	0.81
D	1.97	5.74	0.81

Tabella 9: *Simple and multiple linear regression models together with their prediction performance metrics for the patient-specific dataset.*

Discussion

The problem was framed as an object detection problem and inspiration was taken from some of the most advanced algorithm in the deep learning landscape representing the state of the art, such as YOLO[9]. Detection abilities imply two subtasks: identification of an event and localization of the same along a signal waveform. Despite in these tasks are usually accomplished separately, by first classifying the object and then by understanding its position through a laborious post-processing step, in this thesis project the components have been unify into a powerful, flexible, convolutional neural network optimized for 1D processing. The object detection problem has been reframed as a single hybrid regression-classification problem, straight from signal samples to event coordinates and identification probabilities, reducing computational and time costs.

The overall network architecture can be conceptually divided in two parts with different purposes. The first part, composed by the sequence of convolutional layers interspersed by maxpooling layers, acts as a future extractor, identifying characteristic patterns in data. It is characterized by peculiar elements, such as the presence of 1-sized convolutional layers to reduce the features space from preceding layers, by combining features across all the feature maps. The second part, constituted by the final dense layer providing the output, is essentially responsible for the localization of the identified objects. The output shape is designed so that the input signal segment is divided in five equidimensional windows. If an object falls into a window, that window is responsible for detecting that object. Each window is associate to two output neurons for each label type: one expressing the normalized position of the label inside the segment, the other a confidence scores for this identification, reflecting how confident the model is that the segment contains an object and also how accurate it thinks the segment is that it predicts. The final fully-connected layer can have a variable dimension according to the number of objects to be identified, each one constituting a different output channel C_{output}. Its purpose is to guarantee extreme flexibility about the number of simultaneous detectable labels. Since network performance on each label was comparable when tested on multiple-object detection with respect to single-object case, it means that the features extracted by the first part of the net are general enough to describe entirely the morphological characteristics of signal segment provided as input. Furthermore, it has been shown the possibility to feed the network with a multi-channel input segment, where each channel C_{input} represents a synchronous signal segment. It has been demonstrated also how effective this approach could be: in fact, in

chain configuration, when the network for S1 recognition was fed with detected labels of PPG, the accuracy of prediction, as well as its localization skills improved significantly. This expedient allows to effectively exploit the readily-available information coming from the PPG network to improve the ability to discriminate between S1 and S2 sounds, that could eventually be very similar, especially in noisy signals.

Going into the results of the evaluation on the test set, it can be seen that excellent results were obtained especially for foot and S1 detection. In fact, in the identification task, both the PPG network and the PCG one achieved 97% of balanced accuracy on the cited labels, while on maximum slope detection it was 94%. A definitely worse performance was achieved on PPG peak detection, since balanced accuracy was 85%. A deeper analysis of the statistics showed that the number of false negatives was equal to the number of false positives, thus the network outputs a correct absolute number of identifications even though in a wrong window. The reason of this ability degradation must be sought in the labeling quality. In fact, while in high quality signals the systolic peak is easily identifiable, on real-world data it can be merged with diastolic one. Therefore, the automatic labeling algorithm developed in this project, although quite robust, was not always able to correctly localize the peaks, determining incorrigible noise in the training set that led the network to wrongly locate the labels. Anyway, the robustness of the deep learning algorithm it is evident when the network is fed with reliable samples.

Another characteristic of the network that deserves to be explored is how the segments are processed to perform a detection. Since the network processes the whole segment to perform the detection, this means it uses features from the entire segment as context to predict labels, by reasoning globally about the full segment and all the present objects, accounting also for the encoded presence of contiguous labels.

Another point that requires closer examination is the pre-processing of the input signals the network can work with. In this project have been shown the possibility to work even with almost raw signals, reducing computational cost and time resources needed to perform the detection. The presence of noise in the training data also increases the robustness of the network, acting as a regularization factor, as demonstrated in a noise sensitivity test that was performed.

As for the systolic blood pressure estimation models, as expected the maximum slope label and the foot label were the most useful for BP estimation. Since they were less affected by localization error, more reliable features were extracted with respect to when peak reference point was used. The four most promising models among the tested ones showed an evident linear correlation between estimate and target, with a determination coefficient greater that 0.80, while mean absolute error and mean squared error were less than 2 mmHg and 6 $mmHg^2$, respectively.

Model B, validated in literature with VTT by Esmaili et al. in [5], proposes a reciprocal relationship between SBP and VTT. Here it finds confirmation of its value for systolic blood pressure assessment. In Table 10 are reported their results for the same model accounting also for the different reference points used. It is worth noting the similar determination coefficient for both foot and maximum slope label and the difference in the peak-related one. Very similar results were obtained in this project also with model C, describing a logarithmic relationship between VTT and SBP.

Even better prediction outcomes were obtained by exploiting model A, having a

Label	Study	MAE [mmHg]	R^2
Foot	This work	2.47±2.41	0.75
	[5]	1.14±14.94	0.71
Max slope	This work	1.81±1.36	0.81
	[5]	7.37±11.09	0.84
Peak	This work	2.40±1.15	0.75
	[5]	7.47±11.08	0.84

Tabella 10: *Comparison between results obtained by exploiting model A in this project against the ones obtained by Esmaili et al. in [5].*

coefficient of determination equal to 0.82 and consequently slightly lower estimation error. This model, as model A, describes a reciprocal relationship between VTT and SBP, but here the time delay is measured as an average of all the computed VTTs, namely VTT_{Foot}, $VTT_{Max\ plot}$ and VTT_{Peak}. It can be hypothesized that the mean value computation partially reduces the noise carried by elementary VTTs, reducing the variance of the extracted feature. This model makes the most of the network capabilities, since it simultaneously exploits every object detected by the algorithm.

Conclusions

In this project has been designed and implemented a fast, flexible, accurate neural network for object detection in time series. The same network is able to both recognize and localize multiple events simultaneously with almost the same computational cost. This ability to accomplish different tasks at the same time makes it valid and efficient for different purposes in the most disparate fields and contexts, for example to detect QRS complexes in ECG signals. Furthermore, the possibility to use the same network architecture with a different number of input channels opens up a variety of possibilities and use cases. For instance, the network can be used to process concurrent frequency and time contents. Despite point detection is a simple task, it could be the key to enable several more complex tasks, like, for example, signal segmentation. In this thesis, also a use case demonstration of the network potentialities has been provided. The features extracted by using the network have been successfully employed to estimate the systolic blood pressure, leveraging and strengthening the simple regression models adopted in literature in similar contexts. Although further studies are needed to investigate these data, results are very encouraging and consistent with expectations, as well as literature outcomes.

Indice

Elenco delle figure

Elenco delle tabelle

List of abbreviations

PPG	Photoplethysmogram
PCG	Phonocardiogram
ECG	Electrocardiogram
BCG	Ballistocardiogram
HR	Heart Rate
BP	Blood Pressure
SBP	Systolic Blood Pressure
DBP	Diastolic Blood Pressure
VTT	Vascular Transit Time
PTT	Pulse Transit Time
PAT	Pulse Arrival Time
PEP	Pre-Ejection Period
NN	Neural Network
ANN	Artificial Neural Network
CNN	Convolutional Neural Network

Capitolo 1

Introduction

1.1 Context

Since 2003 the Global Burden of Disease Study, organized by the World Health Organization, has pointed towards hypertension as the most important global risk factor for morbidity and mortality. Hypertension consists in a long-term medical condition in which the blood pressure in the arteries is persistently elevated and it is one of the strongest risk factors for almost all different cardiovascular diseases acquired during life, that are the first cause of death globally.[12] In fact, hypertension has been proven to facilitate coronary disease, left ventricular hypertrophy and valvular heart diseases, cardiac arrhythmias, atrial fibrillation, cerebral stroke and renal failure.[13] An estimated 1.13 billion people worldwide is affected by hypertension, two-thirds of which living in low- and middle-income countries.[14] Furthermore, it is also called "the silent killer" because most people affected by it are unaware of the problem because it may be totally asymptomatic.[14] In 2015, 1 in 4 men and 1 in 5 women had hypertension, but less than 1 in 5 people have the problem under control.[14] From this framework the need to monitor blood pressure level clearly emerges. The technological progress of the last century has provided us many important tools to face and prevent this problem, each with its own advantages and disadvantages.[10] Sphygmomanometers, auscultatory or oscillometric, semiautomatic or automatic, are the preferred method for measuring BP in the doctor's office. Anyway, office BP is often performed improperly, with inadequate attention to the standard conditions recommended for a valid measurement of office BP.[15] Improper measurement of office BP can lead to overestimation of systolic arterial BP and underestimation of diastolic one. In addition, oscillometric and auscultatory BP measurement are particularly sensitive to cuff dimension, resulting in an erroneously high BP reading when the cuff is too small and vice versa, a problem occurring especially in overweight patients. This limits, as well as the lack of adequate oscillometric BP validation studies, should lead to reconsidering the confidence in BP values assessed this way.[15] Furthermore, office BP allows to collect just a limited number of measurements, less representative of daily life. This specific problem is usually solved by adopting non-invasive, out-of-office BP measurement consisting in using either home blood pressure monitoring (HBPM) or ambulatory blood pressure monitoring (ABPM), which provide a larger number of BP measurements than conventional office BP in daily life conditions.[15] Home BP is the average of all BP readings

performed with a semiautomatic, validated, oscillometric BP monitor, for at least 3 days and preferably for 6-7 consecutive days before each clinic visit, with readings in the morning and the evening, taken in a quiet room after 5 min of rest, with the subject seated with their back and arm supported. Instead, ABPM provides the average of BP readings over a defined period, usually 24 hours and it is typically programmed to record BP at 15-30 min intervals, and average BP values are usually provided for daytime, night-time and whole day. Because of their prolonged measurements and the daily life conditions in which those are performed, a major advantage regarding both ABPM and HBPM is that they enable the diagnosis of white-coat and masked hypertension, phenomenons occurring in doctor's office consisting in a overestimation and underestimation of BP, respectively. Additionally, ABPM offers a series of advantages with respect to HBPM, such as a stronger prognostic evidence, night-time readings, measurement in real-life settings, short-term BP variability information.[15] Anyway, these approaches always rely on cuff-based oscillometric technique, that can cause discomfort in the examined subjects and, although ABPM occurs while individuals go about their normal daily activities, they are asked to keep their arm still while the cuff is inflating to avoid excessive motion, which is associated with unobtainable or artefactual readings.[10] In fact, even though other non-invasive, continuous techniques exist, these are impossible to exploit for ABPM because their use is strongly limited to specific application, such as subjects undergoing surgery or in intensive care units.[10] For instance, volume clamp method or radial artery applanation tonometry are both systems providing a continuous arterial BP waveform without the risks of an invasive arterial catheter, but they require to have expensive and complex machinery to be used in a controlled environment.

From this context emerge the desirability of new continuous, non-invasive BP monitoring systems able to guarantee the possibility and the advantages of ABPM, as well as the comfort of cuffless methods.

1.2 Background

A step towards realizing this goal was taken in 2017 by Porro and Vannoni[1][2] from Politecnico di Milano. Their thesis work, raised as a collaboration between Politecnico di Milano and Biocubica S.r.l., led to an operational prototype named Soundi. The device, that should be worn oh the subject's chest in close contact with his skin, is able to continuously record two useful signals: phonocardiographic signal (PCG), through a low-cost pressure sensor, and photoplethysmographic signal (PPG), through an optical system composed by a LED and a receiving photodiode. These signals can be exploited to calculate the beat-to-beat vascular transit time (VTT), a time delay measuring the time spent by the blood pressure wave to transit from the heart to the vascular bed of a generic peripheral site.[4, 6] It is computed as the time difference between a reference point on the PPG, such as the onset of the PPG peak, the maximum positive slope, the systolic peak, and the first cardiac sound of the PCG, occurring at the onset of ventricular systole as effect of the closure of the mitral and tricuspid valves.[3] VTT has been discovered correlated to blood pressure (BP) in several studies [4, 3, 5], potentially enabling the possibility to exploit this device as a cuff-less ABPM instrument.

The physiological principle on which this correlation is based can be modeled in several ways by assuming vessels like elastic tubes and the blood flow in the arteries as the propagation of pressure waves inside them.[2, 3, 5] One of these models is the Moens-Korteweg model, which links the velocity of the blood pulse (PWV) propagating through an arterial tract of length l to its geometric and mechanical parameters as:

$$PWV = \frac{l}{VTT} = \sqrt{\frac{Eh}{2r\rho}}$$

where E, h, r and ρ are the elasticity (Young modulus), the thickness, the inner radius and the density of the vessel, respectively. Since the elasticity can be heuristically modeled as a function of the blood pressure as:

$$E = E_0 e^{aBP}$$

where a is scalar factor, BP can be described as:

$$BP = -\frac{2}{a} \ln VTT + \frac{\ln \frac{2r\rho l^2}{hE_0}}{a} \tag{1.1}$$

where E_0, P_0 and α are individual-specific parameters.

Another widely used model[3, 5], leveraging the same principles, computes the elastic module as:

$$E = E_0 e^{a(BP - P_0)}$$

The compliance C is defined as the rate at which the tube cross section changes as consequence of pressure; considering the conservation of mass and momentum equations:

$$C(BP) = \frac{A_m}{\pi P_1 [1 + (\frac{BP - P_0}{P_1})^2]}$$

where P_0, P_1, and A_m are individual-specific parameters as well. Writing the wave propagation equations inside the vessels leads to the following equation:

$$P(x,t) = f(x \pm t/\sqrt{LC(BP)})$$

where $L = \rho/A$, in which ρ and A are the blood density and vessel cross section, respectively. Therefore, VTT for a tube of length l can be derived as:

$$VTT = \frac{l}{PWV} = l\sqrt{LC(BP)}$$

Furthermore, by substituting L and $C(BP)$, VTT can also be formulated as:

$$VTT = \frac{l}{PWV} = l\sqrt{\frac{\rho A_m}{\pi A P_1 [1 + (\frac{BP - P_0}{P_1})^2]}}$$

which explained with respect to BP finally leads to:

$$BP = P_0 + \sqrt{-P_1^2 + \frac{l^2 \rho A_m}{\pi A} \times \frac{1}{PTT^2}} \tag{1.2}$$

It is worth noting that the cited models all relies on subject-specific parameters, thus in order to be used they require some kind of calibration.

1.3 Proposed work

The objective of this thesis was to exploit the high-quality cardiac signals recorded by Soundi, namely PPG and PCG, to compute vascular transit time in a reliable, computationally-undemanding, robust way with the aim of using it to estimate arterial blood pressure by leveraging the theoretical, physiological basis shown. In order to accomplish the task, the project was divide in three main parts, described in the following.

1. The first part aims at building an automatic algorithm for the identification of all the reference points needed for VTT computation, namely the S1 sound of PCG and the several functionally equivalent points of PPG: the foot, the maximum slope and the systolic peak. To accomplish this task a deep learning approach was chosen, since it allows to process almost-raw signals, considerably reducing the computational costs required by the pre-processing phase. Furthermore, deep neural networks are also potentially more robust to noise, fact that allows to work more effectively with real-world data. This part includes the definition of a protocol for the collection of signals, their collection, the implementation of an automatic labeling algorithm to detect the reference points, the design of a set of suitable neural network architecture to be validated, the definition of datasets to be used for training, tuning and testing the neural networks as well as the training and the evaluation of the same. The problem was framed as an object detection problem and inspiration was taken from some of the most advanced algorithm in the deep learning landscape representing the state of the art, such as YOLO.[9]

2. The second part is dedicated to the crafting of relevant features by exploiting the points detected through the neural network. Here an algorithm for computation of both vascular transit time and heart rate was implemented and evaluated.

3. The third part represents a use case of the algorithms implemented. In this phase were implemented some systolic blood pressure estimation algorithms, which exploit the theoretical, physiological models described before [(1.1), (1.2)] to map the extracted features to the corresponding blood pressure values. Since few studies explicitly evaluate the effectiveness of the different ways to compute VTT for BP prediction, the same models were tested with all the computed VTTs, as well as their combinations obtained with several methods. Also HR was taken into considerations in the models, as done in other studies.[1] This last part comprehends the definition of a protocol for data collection that included the induction of blood pressure changes to cause intra-subject variability, the collection of signals and corresponding BP values, the definition of models to be trained and tested.

Capitolo 2

State of the art

2.1 Non-invasive BP measurement

During the last century, many instruments and methods, exploiting different physical and physiological principles, have been developed in order to measure blood pressure. According to their invasiveness, these methods can be divided in two macro categories: invasive instruments, that perform direct measures, and non-invasive ones, that estimate BP leveraging other principles. In this chapter the theoretical basis of the project are further explained and an overview about the existent BP measurement approaches is presented, with a particular focus on non-invasive methods and deep learning approaches.

Invasive blood pressure measurements The instruments for invasive blood pressure measurement quantify the blood pressure directly in the vessel. Although many different instruments exist, all of them are based on a catheter, that is inserted in an artery, usually in the brachial or in the aortic one.[10] This invasive arterial line constitutes the gold standard for blood pressure measurement for critically ill patients and for BP monitoring during higher-risk surgical or interventional procedures.[10, 16] This method increases patient safety in the ICU or during lengthy surgical procedures by allowing regular blood sampling for laboratory testing and blood gas analysis.[10] Furthermore, since it measures directly the pressure of the circulatory system, it is also the most accurate, as well as the only one able to provide directly continuous measurement when the instrument is correctly placed and calibrated.[10] Anyway, this widely accepted clinical practice is affected also by weaknesses, risks and pitfalls.[10] The placement of arterial catheters is associated with rare but potentially severe complications, such as permanent ischaemic damage, pseudoaneurysm, infection of the insertion site, catheter-related blood stream infections and haematoma/bleeding.[10, 16] Although these severe complications are rare, they become relevant in regard to the millions of arterial catheters that are placed perioperatively and in the ICUs each year, resulting in a significant cost of health care.[10] Therefore, when comparing new BP measurement technology with the invasive arterial catheter, we must also consider the above-mentioned potential pitfalls of this clinical gold standard method.[10]

Method	Description	Advantages	Limitations
Arterial line	Direct arterial catheterization connected through a fluid column to a pressure transducer	Gold standard for invasive monitoring Continuous monitoring	Risk to the patient Requires trained personnel to place appropriately
Korotkoff	Manual auscultatory technique where cuff is inflated and slowly deflate while practitioner listens over brachial artery for Korotkoff sounds	Clinical gold standard for non-invasive monitoring Inexpensive No risk to patient	Relies on trained personnel Subjective method with possible human error Cuff sizing error Cannot be done continuously
Oscillometric	Automated technique where cuff is inflated and slowly deflated while a pressure sensor detects oscillations	Convenient Inexpensive Requires little to no operator skill	Cuff sizing error Cannot be done continuously
Volume clamp	Method which utilizes a finger cuff that adjusts pressure order to keep optically measured finger vascular volume constant	Convenient Non-invasive Continuous monitoring	Expensive Affected by factors such as cold extremities, vasopressors Needs calibration via another method
Tonometry	Uses applanation of the radial artery and measures pressure transmitted through skin	Non-invasive Continuous monitoring	Expensive Manual positioning of tonometer over radial artery not always accurate
Pulse transit time (photometric)	Pulse transducers located at two different sites records the time it takes for a arterial wave to travel between these two points	Non-invasive Continuous monitoring Requires no additional monitors over standard ASA monitors	Not fully developed to correlate pulse transit time to blood pressure

Tabella 2.1: *Different techniques for blood pressure monitoring along with their description, main advantages and main disadvantages.*

Non-invasive methods estimate the blood pressure by exploiting some physiological events that are correlated with the blood pressure itself. Although a complete replacement of continuous invasive BP monitoring by continuous non-invasive technologies in ICU and high-risk surgery patients is currently not possible because of the need for regular arterial blood gas analyses, non-invasive technologies should be evaluated as a potential alternative in other frameworks.[10] In particular, especially in the ABPM framework non-invasive methods constitute an added value because of their comfort with respect to intermittent oscillometric BP measurements and invasive arterial catheter-derived BP measurements.[10] In the next sections, an overview of the most validated and consolidated instruments and technologies for BP estimation will be presented, followed by some other experimental, promising techniques.

2.1.1 Sphygmomanometry

In 1896 Riva-Rocci introduced the upper-arm cuff for sphygmomanometry.[10] Riva-Rocci's blood pressure cuff was a simple device consisting of a rubber bag, surrounded by an inelastic material, which could be inflated by air with an attached rubber bulb along with a manometer to measure pressure.[16] Years later, an understanding of the pulsatile nature of the systemic arterial blood stream, combined with previous efforts at estimating BP using primitive devices, allowed Korotkoff to develop the auscultatory technique.[17] He understood that the application of external pressure exceeding systolic BP would completely pause blood flow and that as this external pressure was gradually reduced, blood flow would return in a predictable way. Korotkoff clearly outlined the different sounds which could be heard when taking measurements by deflating an air-inflatable cuff.[17]

During the last century this method has changed very little and it continues to be a simple and inexpensive way of obtaining blood pressure.[16] The cuff is placed on the proximal arm, inflated to a pressure estimated to be above systolic pressure and deflated slowly while a stethoscope is placed over the brachial artery to listen Korotkoff sounds, which consist of five distinct phases.[16] Phase I, manifested as first clear tapping sounds, corresponds with systolic blood pressure. In phase II and III, that have no clinical significance, sounds become firstly softer and longer and then crisper and louder. Phase IV is generally described as muffled sounds which disappear once phase V is reached. The beginning of phase V corresponds with diastolic pressure, but this transition may be difficult to determine.[16] Mean blood pressure (MAP) is calculated by the following equation[16]:

$$MAP = \frac{2}{3}DBP + \frac{1}{3}SBP$$

Originally, blood pressure devices were mercury sphygmomanometers but, due to concerns of mercury exposure, aneroid and hybrid devices have become increasingly popular.[16] Mercury sphygmomanometers use columns of mercury to display cuff pressure while aneroid devices use a system of metal bellows and a circular scale to display pressures. Hybrid devices use instead an electronic pressure gauge and display values digitally. Anyway, these devices require more maintenance and should be calibrated to a mercury sphygmomanometer every 6 months.[16] If used appropriately, both devices

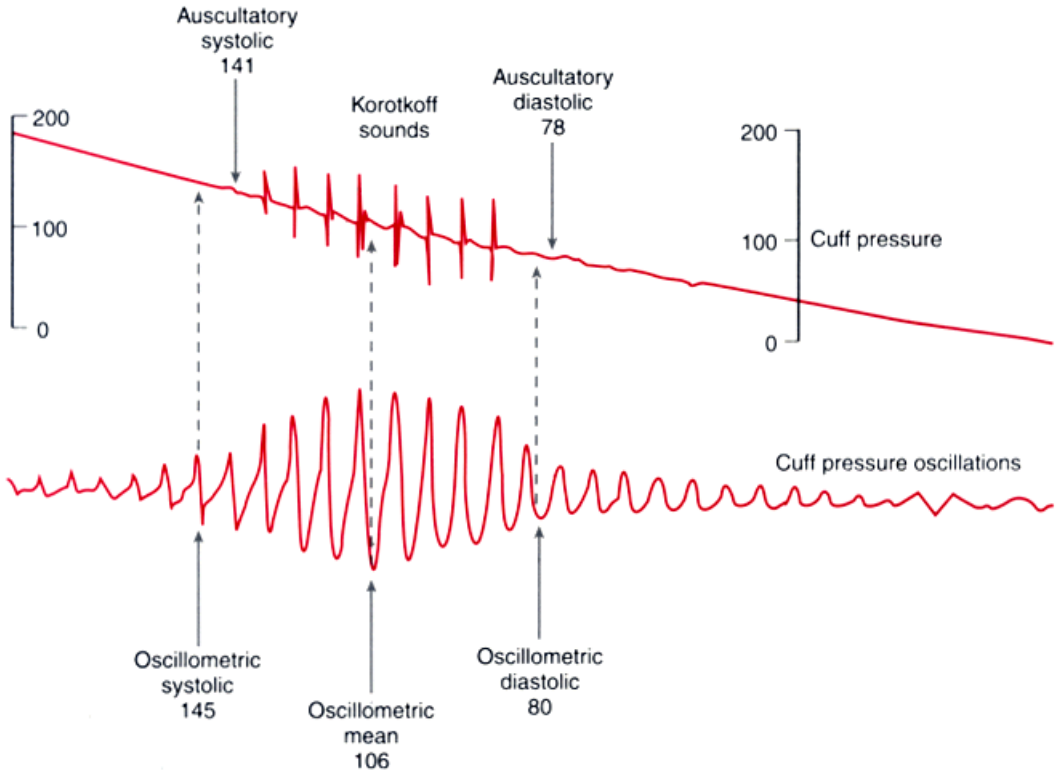

Figura 2.1: *Comparison of blood pressure measurements by Korotkoff sounds and oscillometry. Oscillometric systolic blood pressure is recorded at the point where cuff pressure oscillations begin to increase, mean pressure corresponds to the point of maximal oscillations, and diastolic pressure is measured when the oscillations become attenuated. Note the correspondence between these measurements and the Korotkoff sounds that determine auscultatory systolic and diastolic pressure. (Redrawn from Geddes LA: Cardiovascular Devices and Their Applications)*

have been shown to be comparable to a mercury sphygmomanometer with the added benefit of being easier to use.[16]

Anyway, this method has also several disadvantages and weaknesses. In order to report reliable measurements, cuffs should respect some crucial requirements, such as the width, that should be 46% of the arm circumference, and the placement.[16] Even if optimal conditions occur, Riva-Rocci-Korotkoff measurements underestimate systolic blood pressure and overestimate diastolic blood pressure when compared to invasive catheter measurements.[16] Furthermore, it requires enough blood flow to generate the turbulence needed to produce Korotkoff sounds.[17] In instances in which blood flow to an extremity may be compromised, this technique may fail due to the inability to detect Korotkoff sounds.[17] Also, sphygmomanometry requires that cuff inflation lead to arterial occlusion and consequential discomfort.[17] In extremely large patients, or in those with excessively calcified arteries, this occlusion may also not be possible.[17] Unfortunately, although quite simple to learn, sphygmomanometry is also time consuming and labor intensive.[16, 17] Since it is difficult to automate, semi-continuous measurement and its adoption during anesthesia are impractical.[17] Currently, automated blood pressure cuffs are ubiquitous in operating rooms across the world, but these are based on a different principle: the oscillometric technique.[16, 17]

2.1.2 Oscillometric method

The principle of oscillometry was first described by Marey in 1875.[10, 16, 17] As in sphygmomanometry, oscillations of the arterial wall are detected by an upper-arm cuff, but, unlike the auscultatory technique, in which the return of blood flow is detected audibly, the oscillometric technique measures the changes in cuff pressure that occur when blood flow returns during deflation.[17] When the cuff pressure exceeds systolic BP, small-amplitude oscillations occur and can be detected. While cuff pressure is decreasing, the amplitudes increase and reach their maximum when the cuff pressure level corresponds to the mean BP.[10, 16, 17] Then, the arterial wall oscillations decrease below a cuff pressure equivalent to the diastolic BP resulting in a continuous reduction of the amplitudes. In this framework, maximum systolic BP corresponds with the cuff pressure when the initial increase in arterial oscillations happens, while diastolic BP corresponds with the lowest cuff pressure just before oscillations stop decreasing in amplitude.[10, 16, 17]

However, oscillometry must detect small pressure changes within the cuff despite a relatively rapid reduction in cuff inflation pressure. Furthermore, since a standard algorithm for detecting the oscillometric systolic or diastolic BP is not available, multiple proprietary algorithms exist and for this reason they should not be considered as reliable as MAP.[10, 16, 17] Because the physical principles on which they are based are similar, the oscillometric technique suffers some of the same disadvantages as the auscultatory technique, including poor performance in larger patients and patients with calcified arteries, as well as in those with reduced blood flow, intermittent measurements, cuff sizing sensitivity.[10, 16, 17] One of the largest comparisons of invasive against non-invasive BP measurements was made by Wax et al,[18] who retrospectively compared BP from automated oscillometric readings with arterial BP readings derived from arterial line in more than 15,000 non-cardiac anesthetics. They found

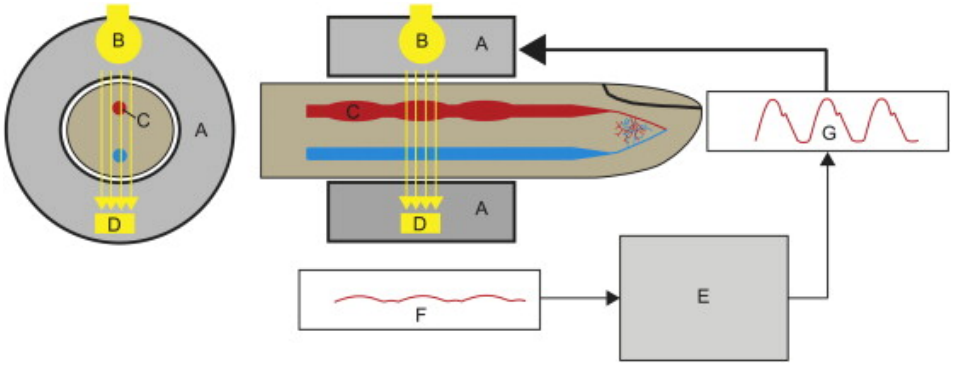

Figura 2.2: *Schematic illustration showing the theoretical principle of the volume clamp method (also called vascular unloading technique). An inflatable finger cuff (A) applies pressure to the finger and contains an infrared transmission plethysmograph (B) to measure the finger artery's (C) diameter (i.e., blood volume). A light detector (D) that is also integrated in the finger cuff measures the absorption of the infrared light. An increase in artery size because of an increase in blood volume (and subsequently pressure) automatically leads to an increase in cuff pressure with the help of a photo-plethysmographic control system (E) in order to keep the artery diameter constant (and the arterial wall 'unloaded'). Thus, from the pressure needed to keep the volume in the finger artery constant (F) throughout the cardiac cycle, the arterial blood pressure waveform can be derived indirectly (G). (Redrawn from [10])*

the standard deviation of the difference between the two to be approximately ±12.5 mmHg when MAP corresponds to 75 mmHg,the median. At the extremes, namely hypertension or hypotension, the standard deviation worsened substantially.[18]

Today, automatic oscillometric BP measurement devices have almost completely replaced the auscultatory method,[10] since it allows to avoid the inconvenience of auscultation and to automatize the process. The oscillometric method is broadly and confidently used in physicians' offices, emergency departments, hospital wards, for perioperative monitoring and treatment in the majority of general anesthesia cases.[10]

2.1.3 Volume clamp technique

In 1973 the Czech physiologist Jan Penaz described a new method for non-invasive BP measurement called "volume clamp method" and also know as "vascular unloading technique". This approach use finger arteries to quantify BP and has the advantage to be non-invasive, avoiding the risks related to arterial catheterization.[10, 16, 17] The basic device consists of an inflatable finger cuff containing an infrared transmission plethysmograph: a light source is positioned on one side of the cuff and an infrared receiver on the opposite side. This configuration has the ability to estimate the finger artery's diameter and the blood volume of the finger through the infrared light absorbance. An increase in blood volume and in artery size is generally associated with an increment in BP and consequently in an increase in absorbed light. The PPG signal is used in a feedback loop to adjust the cuff pressure in order to keep the artery diameter constant. Thus, the BP can be derived from the cuff pressure that is required to keep the vessels in a stationary state of vascular unloading.[10, 16, 17]

However, this technique has also some limitations. The intensity of the BP signal and consequently the reliability of the measurement can be affected by certain clini-

cal conditions, such as finger edema, impaired peripheral blood perfusion or marked hypothermia, that make BP recording difficult or even impossible. Because of these conditions mainly occur in critically ill patients, volume clamp method is made ineffective in 2-17% of patients, depending on the device used.[10] When non-invasive continuous BP monitoring is preferred, it is important to take into consideration these contraindications to properly select the patients that benefits the most of this method. Opportune applications include the continuously monitoring BP in acutely ill patients on hospital admission, during general anesthesia or during the first days after larger surgical interventions.[10, 16, 17] There have been multiple comparisons of volume clamp devices with direct arterial catheterization. Overall, data suggest performance that is at least comparable to oscillometric techniques.[17]

Since Penaz, volume-clamp technique has improved and today different non-invasive BP measurement technologies based on the volume clamp method are available. Wesseling combined the volume clamp technique with Penaz' feedback loop, producing the Finapres (an acronym for FINger Arterial PRESsure device).[10, 16, 17] The ClearSight device uses a physiologic calibration algorithm to achieve vascular unloading. It induces periodic, stepwise bladder pressure changes and simultaneously reanalyzes the PPG waveforms. If alternate pressure settings result in larger PPG amplitudes, the feedback set point is adjusted to ensure a maximally unloaded state of the finger artery.[17] The Continuous non-invasive Arterial Pressure (CNAP) device uses the "Vasomotoric Elimination and Reconstructed IdentiFication of the Initial set point" (VERIFI) algorithm to unload peripheral arteries, coupled with an upper arm oscillometric BP monitor for calibration. The CNAP is designed to account for the fact that no feedback mechanism can be completely instantaneous. To generate an initial BP measurement and identify the set point, the VERIFI algorithm relies on a calibration done with brachial artery cuff.[17]

2.1.4 Tonometry

O'Rourke et al. developed the field of arterial tonometry.[17] The applanation tonometry required the flattening of superficial artery (e.g., radial) underneath a sensor supported by a bony structure (e.g., the styloid bone), allowing a continuous non-invasive beat-to-beat recording of the arterial BP waveform.[10, 16, 17] In an effort to develop an automated tonometer, Tensys Medical developed the T-line, a tonometer stabilized over the radial artery of the wrist by a locking clamp. Once this clamp is affixed over the radial artery, it can periodically adjust the tonometric pressure sensor by locating the ideal measuring point.[17] The T-Line system consists of the following main components: a disposable wrist splint for optimal positioning, a bracelet including the sensor that is applied over the distal radial artery pulse and the sensor positioning mechanisms, and a patient monitor for BP waveform display. The bracelet uses two motors to move the pressure sensor laterally to identify the site of maximum pulsation and axially to achieve optimal artery applanation. As in the oscillometric BP measurements, the MAP, corresponding to the maximum pulse pressure, can be obtained when the transmural pressure is zero. The proprietary signal processing algorithms to measure arterial BP are based on subject specific parameters, such as the body mass

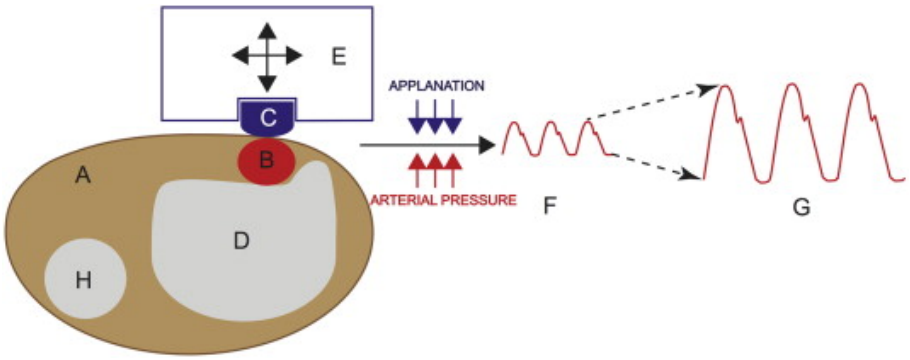

Figura 2.3: *Cross section of the wrist (A) with a schematic illustration showing the radial artery applanation tonometry technology. The applanation tonometry requires that the radial artery (B) be flattened (applanated) underneath a sensor (C) and be supported by a bony structure (e.g., the radius (D)). The sensor is integrated in a bracelet (E) that also holds the motors to move the pressure sensor laterally for identifying the site of maximum pulsation and up and down for achieving the optimal artery applanation. The maximum pulse pressure (equivalent to the mean blood pressure) can be obtained when the transmural pressure is zero (representing the maximum compliance of the arterial wall). Continuous non-invasive beat-to-beat recording of the arterial blood pressure waveform can thereby be derived (F). Subsequently, the arterial blood pressure waveform is scaled using proprietary signal processing algorithms based on biometric data (G). The ulnar bone is indicated by H. (Redrawn from [10])*

index. Estimation is done leveraging a large invasive radial artery reference database and don't need external calibration.[10]

Radial artery applanation tonometry is affected by some specific limitations, such as the need for adequate training of the operator in handling the T-Line system and high sensitivity to motion. Accounting for this pitfalls is crucial to select the right target patient population that will have the maximum benefit, for example patients undergoing elective surgical procedures who are usually monitored using intermittent oscillometric measurements.[10] Several studies have compared invasive arterial measurements of mean arterial BP with estimates from a tonometer during general anesthesia and surgery. The median standard deviation of the difference between the two is 6.3 mmHg.[17]

2.2 Non-invasive methods based on PPG signal processing

Besides the previously explained, widely used devices and techniques, in the last years several studies have shown the potential of other non-invasive methods, based on the processing of either single signal such as ECG, PPG, PCG and BCG or combination of them. [6] In particular, PPG, an optical measurement technique that can be used to detect blood volume changes in the microvascular bed of tissue, has received the greatest attention, because of important technological advancement and numerous confirmation of its potentiality for BP non-invasive estimation. [6, 19]

PPG technology requires only a few opto-electronic components: a light source to illuminate the skin and a photodetector to measure the small variations in light

intensity due to absorption changes associated with modifications in perfusion in the catchment volume.[19] Since the maximum pulsatile component of reflected light occurs approximately in the range between 510 and 590 nm, the green (565 nm) or yellow (590 nm) light is generally used for reflective PPG sensors. However, the red (680 nm) or near-infrared (810 nm) light is generally used for transmissive PPG devices, with the infrared light having the deepest penetration. Because of the optical absorption of hemoglobin is a function of oxygenation and optical wavelength, the use of PPG at multiple wavelengths is also routinely used in pulseoximetry.[6] Since the amount of optical absorption or reflection depends on the amount of blood that is present in the optical path, the PPG signal is responsive to changes in the volume of the blood. The sensor coverage area includes both veins and arteries, and numerous capillaries, hence the PPG signal is a complex mixture of the blood flows in cardiovascular circulatory system.[6] A raw PPG signal generally includes two components, namely a pulsatile and a non-pulsatile blood volume. The pulsatile component of the PPG waveform is often called the 'AC' component and usually has its fundamental frequency, typically around 1 Hz, depending on heart rate. This AC component is superimposed to a large quasi-DC component that varies slowly due to respiration, vasomotor activity and vasoconstriction waves, Traube Hering Mayer waves and also thermoregulation. These characteristics are body site dependent. With suitable electronic filtering and amplification both the AC and DC can be extracted for subsequent pulse wave analysis.[19]

The appearance of the PPG pulse is commonly divided into two phases: the anacrotic phase, the rising edge of the pulse that is primarily concerned with systole, and the catacrotic phase, the falling edge of the pulse that is related with diastole and wave reflections from the periphery. In literature, many characteristic features have been identified and described:[20]

- Systolic amplitude: it is an indicator of the pulsatile changes in blood volume caused by arterial blood flow around the measurement site. It is related to stroke volume, local vascular distensibility and blood pressure.
- Pulse width: it is the pulse width at the half height of the systolic peak. It correlates with the systemic vascular resistance.
- Pulse area: it is the total area under the PPG curve. This area can be divided into two subsections split by the dicrotic notch, whose ratio is correlated to total peripheral resistance.
- Peak to peak interval: it is the distance between two consecutive systolic peaks. It is closely correlated with the RR interval in the ECG signal and widely used to detect the heart rate.
- Pulse interval: it is the distance between the beginning and the end of the PPG waveform. It has been demonstrated highly correlated to HRV in ECG.
- Augmentation index: it is the ratio between the height of the late systolic peak and the early diastolic peak in the pulse. It measure of the contribution that the wave reflection makes to the systolic arterial pressure and it is correlated to a reduced compliance of the elastic arteries.

- Large artery stiffness index: it is measured as the ratio between the subject height and the distance between systolic and diastolic peak. It has been shown correlated to large artery stiffness and age.

Together with the PPG signal itself, PPG derivatives can be found in literature. The first derivative is usually associated to the detection of the diastolic peak, the crest time, and the interval between the systolic peak and the diastolic peak. While the first derivative is hardly used, the second derivative has a wide range of applications due to the fact it is an indicator of the acceleration of blood in vessels.[20]

PPG, despite its simplicity, has been widely employed in the biomedical field because of its wealth of carried information. Main applications are:[19]

- Clinical physiological monitoring, in particular of blood oxygen saturation, heart rate, blood pressure, cardiac output and respiration;
- Vascular assessment, especially for diagnosis of arterial diseases and vasospastic conditions, estimation of arterial compliance and aging, assessment of endothelial function, microvascular blood flow and tissue viability, and venous assessment;
- Study of autonomic functioning, like vasomotor function and thermoregulation, blood pressure and heart rate variability, orthostasis, neurological assessment.

As previously mentioned, PPG can be very useful and effective for assessment of BP in subjects and many techniques with this aim have been developed and validated. These methods, that exploit the known information carried by this biosignal, are widely used even if the characteristics of the PPG waveform have not yet been fully understood.[6] These approaches can either rely just on the PPG signal or leverage other signals in combination with the PPG itself. In the following paragraphs will be shown how, in particular, PPG can be useful for BP estimation, alone or coupled with other time series.

2.2.1 PPG only

BP estimation by using only PPG signals can be performed with two approaches: the first one implies the analysis of the wave form and its derivatives to extract relevant geometrical and dynamical features correlated with vascular pressure, the second one uses multi-source PPG in order to asses time differences between corresponding points in the same wave at distant body locations.

PPG features The idea of estimating BP from a single PPG signal has been investigated in several studies. Many authors have reported a strong correlation between BP and features obtained from the PPG signal. These features can be either hand-crafted or automatically learned by machine learning algorithms and can be extracted from either the PPG signal or its first and second derivative, which are measures of velocity and acceleration, respectively.

Hand-crafted features are suitable to characterize each cardiac cycle by accounting for different aspects, such as dynamics of systolic, diastolic and dicrotic phases,

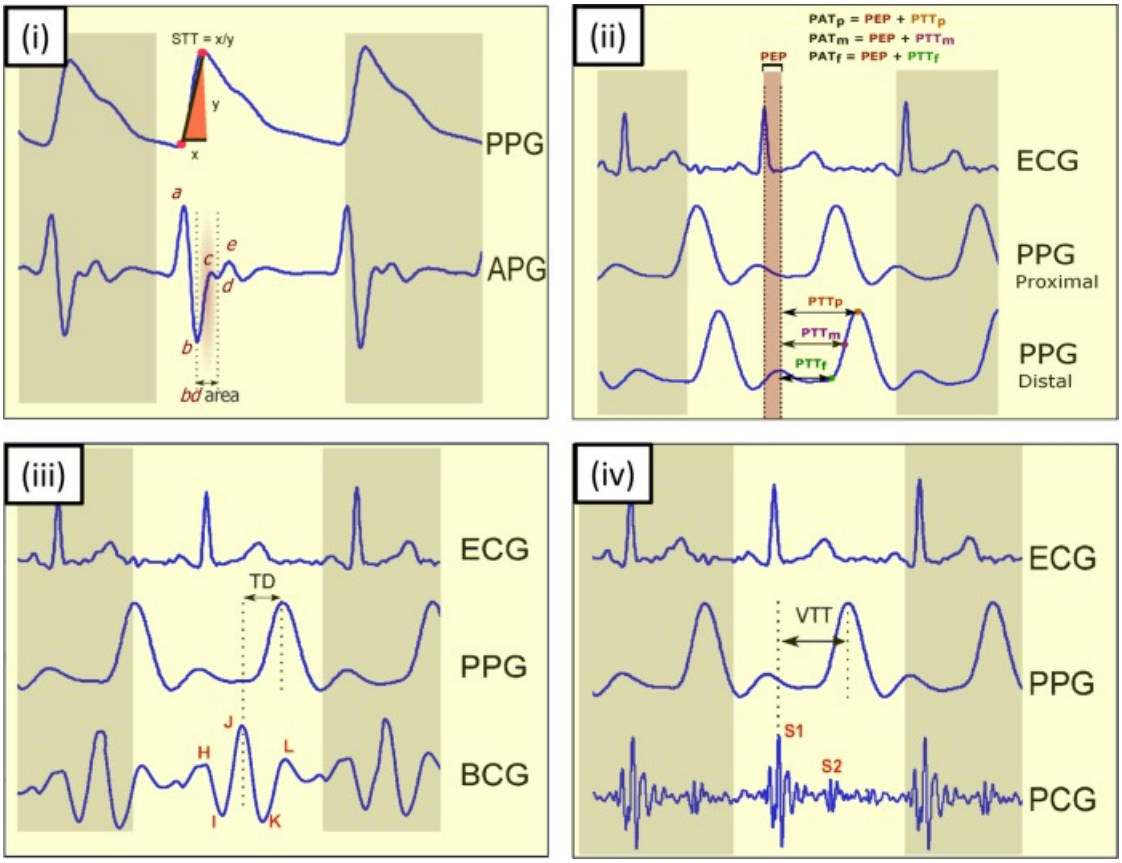

Figura 2.4: *Key features of blood pressure estimation using PPG and other physiological signals. (i) Using PPG signal and its derivative, (ii) using ECG and PPG signals, (iii) using BCG signals and PPG signals, and (iv) using PCG and PPG signals. Here, PPG photoplethysmogram, APG acceleration photoplethysmogram, BCG ballistocardiogram, PCG phonocardiogram, STT slope transit time, PTT pulse transit time, PEP pre-ejection period, PAT pulse arrival time, TD time interval between the J peak in the BCG signal and the systolic peak in the PPG signal, VTT vascular time interval between the first heart sound S1 and the systolic peak in the PPG signal, S1 first heart sound, S2 second heart sound. (Redrawn from [6])*

statistical information like the standard deviation, dynamic ratio between cardiac cycles dynamics, stability of cardiac cycles. Some of the proved correlated hand-crafted features are:[6]

- Slope Transit Time. It is a slope parameter calculated from the foot to peak of the systolic waveform. It reflects the steep trend of rising pulse wave.
- bd area. It is the area between the minimum of the second derivative and the following local minimum corresponding to the minimum gradient in the systolic phase. It was demonstrated that the b/a ratio reflects increased arterial stiffness and increases with age.

Besides these validated correlated parameters, many other hand-crafted features have been selected to feed machine learning models, obtaining encouraging results.

In [21], the authors proposed an approach to non-invasive continuous blood pressure estimation based on artificial neural networks (ANN) trained with 21 ad hoc input parameters extracted from a set of 15.000 PPG signals. They extracted the different parameters of every pulse of the PPG signal and uses them as the input of the ANN. Then, the network was trained to provide the corresponding BP value at the output. The employed multi-layer, feed-forward ANN was composed of 21 input neurons (corresponding to the parameters extracted from PPG), two hidden layers and tree output neurons for the prediction of Systolic BP (SBP), Diastolic BP (DBP) and Mean Arterial Pressure (MAP), respectively.

In [22], the authors proposed a method to extract a comprehensive set of 46 features by combining PPG signal-based and heart-rate-variability-related features, using a single PPG sensor. These features were used to feed a combinatorial neural network model and separate models for systolic and diastolic pressures were trained. The ANN regression models were trained with the Levenberg-Marquardt algorithm and then weighted and combined to produce the final output.

In [23], the authors evaluate a method for the estimation of the nocturnal systolic blood pressure dip from 24-hour blood pressure trends using a wrist-worn photoplethysmography sensor and a deep neural network in free-living individuals, comparing the deep neural network to traditional machine learning and non-machine learning baselines. Features based on heart rate variability and pulse morphology were extracted from the PPG waveforms. Long-short term memory networks, dense networks, random forests and linear regression models were trained and evaluated in their capability of tracking trends in BP, as well as the estimation of the SBP dip. Best performance for estimating the SBP dip were obtained with a deep LSTM neural network.

In [24], the authors perform a study whose purpose was to use the characteristics of the pulse wave measured by photoplethysmography to estimate the blood pressure with a multi-dimension regression model by exploiting the assumption that the contour of pulse wave includes some characteristics of the artery.

In [25], the authors propose to estimate the blood pressure from PPG using Multi Task Gaussian Processes and compare it with an artificial neural networks approach. The performance were comparable or better than the existing methods of computing BP from non-invasive data.

In [26] the authors developed a method for estimating systolic and diastolic BP based on the multitaper method for spectral feature extraction, while an artificial neural network was used for estimation.

In [27], an algorithm was developed using a modified group method of data handling technique to estimate systolic BP and diastolic BP continuously. The estimation did not require to calibration and was done without any knowledge about significant features, and features were selected based on their competency. Four popular regression algorithms, including support vector machine, adaptive boosting, decision tree and random forest, were trained with conventional features based on simultaneous ECG and PPG signals and compared with the proposed system. The promising, proposed algorithm was not dependent on synchronization of ECG and PPG.

In [28], the authors proposed a deep-learning-based system that relied on a comprehensive dataset of features crafted from PPG. These features fed a combination of two ANN: the first one was a Multi-Layer Perceptron with a modified Polak-Ribiere back-propagation learning algorithm, the second one an advanced modified version of the Self-Organizing Map motor map.

In [29], the authors developed a data-driven BP estimation, by combining discrete wavelet transform for feature extraction, forward feature selection and support vector machines with RBF kernel for estimation.

Very recently new studies have shown the possibility to apply deep learning techniques for BP estimation without explicitly computing features, but using raw or filtered signals segments as input for convolutional and recurrent neural networks. For example, in [30], the authors the PPG alongside its first and second derivative as inputs into a spectro-temporal deep neural network with residual connections. In a leave-one-subject-out experiment on MIMIC dataset the network was able to model the dependency between PPG and BP, achieving mean absolute errors of 9.43 for systolic and 6.88 for diastolic BP. Additionally they shown the importance of calibration of models on patients, underling that deriving a good general predictive model is difficult. One paramount issue of artificial neural networks is that it is hard to understand how they map input features to output targets because of their intrinsic complexity and functioning. This way, identifying which parameters are the most correlated to BP, without fully comprehending how features are learned, selected, combined and weighted, is a complex reverse-engineering problem.

Pulse Wave Velocity Pulse wave velocity (PWV) is the velocity at which the blood pressure pulse propagates through the circulatory system and can be readily measured non-invasively. PWV is used clinically as a measure of arterial stiffness and arterial distensibility, thus is related to BP. It can be computed as the ratio between a distance between two vascular sites and a time delay in blood propagation.[6, 7]

$$PWV = \frac{Distance}{Time\,Delay}$$

This time delay can be estimated in several ways, according to the type of signals measured at the two sites. If only PPG signals are used, this time interval is named Pulse Transit Time (PTT). It refers to the time taken by a pressure wave to travel between two arterial sites.[6, 7] In order to measure it, two optical systems are placed

at a fixed distance, calculating the temporal shift of corresponding points in two waves. Ears, toes and fingers are common sites used for measurement. PTT has been confirmed negatively correlated with blood pressure and several mathematical models for prediction, both linear and non-linear, have been designed and tested. Some of them are shown in Table 2.2.[6, 7]

2.2.2 PPG coupled with ECG

Another popular method to measure the time delay is based on the time difference between the R-peak of ECG and a characteristic point of PPG peak. Different point of reference on the PPG waveform, such as foot, peak and maximum gradient of the rising edge (the maximum of the first derivative), have been considered to estimate this delay, named pulse arrival time (PAT). Though PAT and PTT parameters are often used interchangeably, they are defined differently. Indeed, PAT includes both the PTT and the pre-ejection period (PEP), which is the time delay between the electrical depolarization of the left ventricle and the start of the mechanical ventricular ejection.[6, 7]

$$PAT = PPT + PEP$$

PEP depends upon a number of factors related to the preload, HR and contractility of the heart. Therefore, the inclusion of PEP in the time computation of pulse wave transmission along the vessel can then induce inaccuracy due to these factors.[4] Furthermore, the impact of PEP on the overall PAT decreases with distance from the heart, thus for near measurement locations it can not be neglected and should be taken into account. The effect of including PEP in BP estimation and the correlation of PAT to BP with respect to PTT are still under investigation.[7]

In Table 2.2 are reported some models used in literature to approximate the relationship between PAT and BP. Although it is proven that PTT has a strong correlation with the systolic and diastolic blood pressures, this relation is highly dependent to each individuals physiological properties. For this reason the need for individual short-term or long-term calibration has been confirmed by a body of literature as a means of improving the accuracy of BP estimation. Despite this, in [40] the authors proposed a novel method that was calibration-free. They extracted several physiological parameters from PPG, such as PTT with respect to R-peak of ECG, HR, Augmentation Index, Large Artery Stiffness Index, Inflection Point Area ratio, to be used as inputs for Regularized Linear Regression, Artificial Neural Networks and the kernelized Support Vector Machines. The results show that the accuracy of the proposed method achieves grade B for the estimation of the diastolic blood pressure and grade C for the estimation of the mean arterial pressure under the standard British Hypertension Society protocol. Successively, in [41], the same authors estimated BP from roughly a thousand subjects from the MIMIC II database. By using PTT-related features, several regression models were evaluated in a 10-fold cross validation experiment. The best results were achieved using Adaptive Boosting (AdaBoost), specifically the mean absolute error (MAE) of 11.17 for SBP and 5.35 for DBP. They also used person-specific calibration in their work, which improved the results.[30]

Mathematical model	Tested Time Delays	References
$BP = a\,ln(\,Time\,Delay\,) + b$	PTT, PAT, VTT	[31, 1]
$BP = a\,(\,Time\,Delay\,) + b$	PTT, PAT, VTT	[32, 33, 34, 35, 36]
$BP = \frac{a}{(\,Time\,Delay\,)^2} + b$	PAT	[37]
$BP = b + \left(\frac{a}{(\,Time\,Delay\,) - c}\right)^2$	PAT	[35]
$BP = \frac{a}{(\,Time\,Delay\,) + b}$	PAT	[38, 39]
$BP = b + \sqrt{c + a\frac{1}{(\,Time\,Delay\,)^2}}$	PAT, VTT	[3]
$BP = \frac{a}{(\,Time\,Delay\,)} + b$	VTT	[5]

Tabella 2.2: *In the first column are reported literature models for BP estimation, in the second one time delays used to test the models, in column three references to studies that proposed the models.*

It is worth noting that, as for PPG-only approach, besides canonical mathematical models several machine learning approaches have been tested also for PPG-ECG approach.

In [42], the authors evaluated the correlation of PPG and ECG signals as well as biometric parameters such as weight, height, body mass index, length and circumference of the arm with BP. Then, they selected three parameters among the others (PTT, weight, arm length) according to this statistical analysis. With these parameters as inputs, they compared artificial neural network with multiple linear and non-linear regressions as an estimating method of BP, with the performance of a KEDA-approved device as reference. The results showed that the ANN had better accuracy than the multiple regression, satisfying AAMI standard as a BP device.

Existing methods for arterial BP estimation directly map the input physiological signals to output BP values without explicitly modeling the underlying temporal dependencies in BP dynamics. As a result, these models suffer from accuracy decay over a long time and thus require frequent calibration. In [43], addressed this issue by formulating BP estimation as a sequence prediction problem in which both the input and target were temporal sequences by using deep recurrent neural network, in particular multilayered, bidirectional LSTM networks with residual connections. The proposed deep RNN model surpassed the accuracy of traditional BP prediction models. The experimental results suggest that modeling the temporal dependencies in BP dynamics significantly improves the long-term BP prediction accuracy.

The majority of studies that predict BP rely on additional efforts to manually extract features from various biomedical signals and, as mentioned above, individual periodic calibration. In order to overcome these limitations other deep learning approaches have been experimented.

In [44], authors proposed a CNN to estimate the SBP using both ECG and PPG signals. Two different methods were investigated and compared: the first method was based on continuous wavelet transform (CWT) and CNN, the second method on random sampling within the stochastic gradient descent (SGD) optimization of CNN fed with raw ECG and PPG. Since both methods are capable of learning how to extract relevant features from the signals, they concluded there is no need for engineered feature extraction.

In [45], the authors proposed a cuffless BP prediction method based on a deep convolutional neural network that can use raw signals for training to predict BP accurately without calibration. They proposed two schemes: extraction through multiple dilated convolution, and concentration through strided convolution with a large kernel, to process sequential ECG and PPG signals through CNN. They evaluated the model with and without calibration, with either time sequences or frequency content and with both, with only PPG signals or PCG signal and both. They obtained the best performance when both ECG and PPG signals were used together, with time content only and, predictably, including a calibration step. The proposed method achieved excellent performance in predicting both systolic blood pressure and diastolic blood pressure fulfilling international standard protocols, AAMI, and BHS.

2.2.3 PPG coupled with PCG

During the systolic and the diastolic phase of the cardiac cycle, audible sounds are produced from the opening and the closing of the heart valves, the flow of blood in the heart, and the vibration of heart muscles. Usually, four heart sounds are generated in a cardiac cycle. Phonocardiography (PCG) is simply the method for obtaining recordings of these sounds by auscultation. The first heart sound and the second heart sound can be easily heard in a normal heart through a stethoscope placed on a proper area on the chest. The normal third and the fourth heart sound are seldom audible in normal individuals through the conventional mechanical stethoscopes but can be detected by sensors with high sensitivity, such as electronic stethoscopes and phonocardiography systems. Sounds other than these four, called murmurs, are abnormal sounds resulting from valve problems, or sounds made by artificial pacemakers or prosthetic valves. The first heart sound (S1) occurs at the onset of ventricular systole. It is characterized by higher amplitude and longer duration in comparison with other heart sounds. It result from the closure of the mitral and tricuspid valves and the vibrations set up before aortic ejection. Its frequency components lie in the range of 10-200 Hz. The second heart sound (S2) occurs within a short period once the ventricular diastole starts. It coincides with the completion of the T-wave of the electrocardiogram (ECG). S2 consists of two high-frequency components, one because of the closure of the aortic valve and the other because of the closure of the pulmonary valve. The second heart sound usually has higher-frequency components as compared with the first heart sound. Many heart diseases are associated wit a characteristic changes of S2. S1 and S2 are basically the main two heart sounds that were used for most of the clinical assessment based on the phonocardiography auscultation procedure.[46]

Some studies show that is possible to estimate BP by performing a heart sound analysis, both in frequency and in time domain. In a study, phonocardiography (PCG) was measured with an esophageal stethoscope: the ratio of S1 to S2 displayed the highest correlation with the systolic blood pressure, with both S1 and S2 amplitude positively correlated to it.[47] Encouraging results were also obtained from the frequency analysis of the S2 sound: time, frequency, and wavelet-based features were extracted and associated to BP with a multi-regression model[48]. In [49], authors extracted features like heart rate, S1 and S2 duration, diastole and systole duration, S1 and S2 growth constants, S1 and S2 growth initial values, S1 and S2 decay constants, S1 and S2 decay initial values, S1 and S2 amplitudes and ratio of S1 amplitude to S2 amplitude. These were the considered for the creation of successful predictive semi-empirical models by using multiple regression analyses. Then, the same authors mapped the same features to BP by exploiting the neural network potentialities: good results were obtained by introducing a regularization step accounting for a cuff-based BP measurement.[50]

As previously explained, the contribution of the cardiac pre-ejection period to PTT could be significant, potentially limiting the ability to track BP in a clinical setting. To help limit the influence of PEP, in 2006 Foo et al. [4] introduced the Vascular Transit Time (VTT), defined as the difference between the foot of the PPG waveform and the first PCG heart sound. Their results showed that observed VTT changes were related to systolic BP (R2 = 0.820; p ¡ 0.05), diastolic BP (R2 = 0.517; p ¡ 0.05), mean arterial pressure (R2 = 0.673; p ¡ 0.05). As systolic BP had the strongest correlation, a regression equation was formulated to associate the two parameters obtaining positive

results in prediction. Thus, the use of VTT, which adopts the same working principle as PTT, can be considered as an alternative for such purposes with the advantage that PEP is now not included in all time-related measurements.

Successively, VTT has been exploited for BP estimation also in other studies. In [8], authors implemented an auscultation device that can continuously estimate blood pressure via simultaneous measurement of PCG and PPG signals. Signals were processed by using FIR filter and discrete wavelet transform, then VTT was computed with respect to S1 and PPG peak, together with Ejection Time (ET) and Heart Rate (HR). In the multiple regression model also individual's physiological parameters such as gender, age, height, and weight were included. The results of this study showed that for the range of normal blood pressure, the error in systolic blood pressure is 6.67 $\pm$ 8.47 mmHg, which is very close to the AAMI standard, 5 $\pm$ 8 mmHg, proving the feasibility of the approach.

In [5], the authors derived VTT values from PPG and PCG signals, while a force-sensing resistor (FSR) was placed under the cuff of the BP reference device to mark the moments of measurements accurately via recording instantaneous cuff pressure. For deriving the PTT-BP models, a calibration procedure including a supervised physical exercise is conducted for each individual. The proposed method is evaluated on 24 subjects. For perturbing BP for the calibration procedure, all the subjects underwent a same supervised physical exercise (running about 3 minutes at 8 km/h), able to cause them a discernible increment of BP. Immediately after the exercise, pressure data were collected by measuring several times. Since each subject was sitting during the data collection, the measured BP is generally decreasing through the measurements.The final results proved that when using PCG for VTT measurement with the proposed regression, the BP can be estimated reliably. Since the use of PCG requires a minimal low-cost hardware, the proposed method potentially enables ubiquitous BP estimation in portable healthcare devices. The same working group, in [3], proposed a novel BP estimator nonlinear model, based on the theory of elastic tubes to estimate the BP using VTT/PAT values precisely (Table 2.2). The proposed method was evaluated on 32 subjects. Using the VTT index, the correlation coefficients for SBP and DBP estimation are 0.89 and 0.84, respectively. Using the PAT index, the correlation coefficients for SBP and DBP estimation are 0.95 and 0.84, respectively. The results showed that the proposed method, exploiting the introduced nonlinear model with the use of PAT index or VTT index, provides a reliable estimation of SBP and DBP.

PPG technology has the huge advantage to be low cost, simple, portable and thus suitable for ubiquitous BP estimation. Improvements in digital sensors, signal processing and machine-learning algorithms, as well as physiologic understanding of the cardiovascular system have led to a greater feasibility of pulse waveform analysis for BP assessment. Despite this, it shows some limitations, that constitute the future challenges to be faced. In general, PPG for BP estimation requires several precautions and actions, such as noise elimination, multi-site measurement, multi-photodectors or multi-signal approach, event detection, different models, individual calibration according to the subject's characteristics, calibration drift correction over short-time intervals.[6] Research into the development of such algorithms is still ongoing, but very encouraging results have been obtained.

As shown, when models are based on hand-crafted features, a subject-specific cali-

bration step is required to reliably predict BP. It is worth noting that to find the best fitting curve for VTT/PTT/PAT-based BP estimation, it is essential to vary BP over a wide range in order to have more points to construct the calibration curve. Significant variations in SBP can be achieved using the methods listed in Table 2.3[7], but it is more difficult to vary DBP.

Procedure	Description
Physical Exercise	Graded Bicycle Test, Running, Sit-ups
Posture	Sitting, standing and lying supine
Valsalva Maneuver	Breathing against closed nose/mouth for 30 seconds
Cold Pressure	Placement of ice wrapped in wet cloth on subject's forehead for 2 min, Hand in 4°C water for 1 min
Mental Arithmetic	Counting backwards from 500 in intervals of 7, Continuous addition of 3 digit numbers for 2 min
Relaxation	Slow breathing/Meditative music
Amyl Nitrate	Inhalation of vasodilator
Relaxation	Slow breathing/Meditative music
Anesthesia	Dental anesthesia
Isometric Exercise	Raising legs/arms against pressure
Sustained Handgrip	Clenching one's fist forcefully

Tabella 2.3: *Techniques used in literature to perturb blood pressure.*

Capitolo 3

Materials and Methods

3.1 Materials

3.1.1 Hardware

3.1.1.1 Soundi

In 2017 Porro and Vannoni[1][2] from Politecnico di Milano developed a prototype system exploitable as a cuff-less ABPM instrument. Their work aims at overcoming some technological and functional drawbacks found into date measurement devices. The thesis work, raised as a collaboration between Politecnico di Milano and Biocubica S.r.l., led to an operational prototype, named Soundi, based on an three main technical innovations, subject of a patent pending at European scale (Patent No. EP3248541A1): a) the design of custom system to detect heart sounds by acquiring air pressure signals on the chest interface; b) the on-chip development of the required analog front-end components by leveraging on the FPGA-based μ-controller architecture; c) the optimized hardware implementation featuring low power consumption that allowed continuous signal recording up to 10 hours.

Figura 3.1: *Logo of Biocubica*

In chapter 2 has been explained the BP estimation method based on vascular transit time (VTT), which is the time taken for a blood pulse wave to travel from the heart to a peripheral site of the human cardiovascular system. The device developed in their thesis is able to continuously record the signals needed to calculate VTT beat per beat: PCG signal, through a low-cost pressure sensor, and PPG signal, through an optical system composed by a LED and a receiving photodiode.

Both signals are sampled synchronously at 400Hz by the microprocessor. Acoustic signal are detected by using atmospheric pressure measurements instead of traditional acoustic sensing, which have been shown to be very sensitive to environmental noise. This objective is fulfilled by coupling together the MPXV7002 analog pressure sensor (NXP Semiconductor) and a custom designed amplification chamber, inspired by a bell-shaped stethoscope. Pressure measurement is performed differentially, since the pressure measured by the sensor into the bell is referenced to the local atmo-

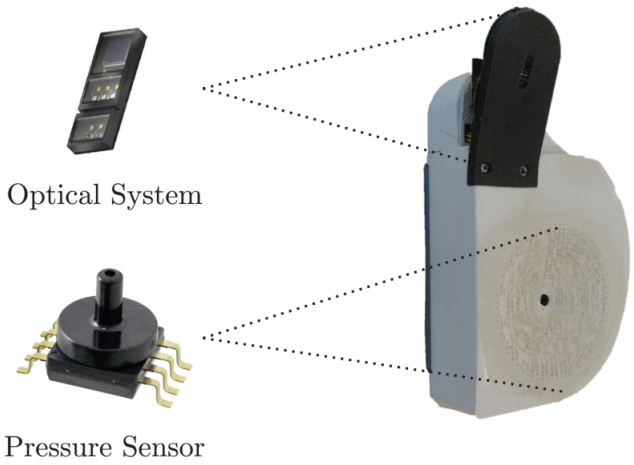

Figura 3.2: *Soundi prototype. Top left: SFH 7060 optical system by BioMon Sensor devoted to PPG detection. Bottom left: MPXV7002 integrated silicon pressure sensor from NXP Semiconductors devoted to PCG detection. Right: front view of the developed device showing the optical sensing area (black) and the pressure sensing area (bell shape). Overall, the device features a height of about 6cm, a width at the base of 4.5cm, a thickness of 1.5cm, and a weight of approximately 100 gr.*

spheric pressure. Optical signal is acquired by the SFH 7060 system (BioMon Sensor - Osram), an analog multi-chip optical sensor embedding a wide-range photodiode (400-1100nm), with good spectral sensitivity (0.5A/W), and three green LED emitters (λ_{PEAK}=530nm), along with one red (λ_{PEAK}=660nm) and one infrared (λ_{PEAK}=950nm) emitters. Soundi integrates both an 8GB microSD card for on-board recording and a wireless communication module implementing a Bluetooth Low Energy (BLE) protocol, useful for real-time signal quality evaluation and for sending commands to the device by means of a custom software running on an host machine. In order to supply both digital (3V) and analog (5V) components by using the 1.5V battery, two boost converters from Texas Instruments are used. The device has to be worn on the left side of the chest, in close contact with the subject's chest skin, and has been realized to meet comfort and ergonomics criteria, even in case of monitoring over a long period.

3.1.1.2 GIMA ABPM50

GIMA ABPM module measures the blood pressure using the oscillometric method with step deflation. It provides two modes of measurement: manual and automatic. Each mode displays the diastolic, systolic and mean blood pressure, together with the HR of the subject. In the manual mode, only one measurement is conducted for each time, while in the auto mode, the measurement is cycled. It is compact and portable, easy to use, suitable for adult, pediatric and neonatal patient. Its clinical accuracy is proved, the BP value of the device is equivalent with the measurement of the stethoscope. The error meets all the conditions in the ANSI/AAMI SP-10:2002+A1:2003 +A2:2006.[51]

	Pressure	Pulse rate
Range	0-290 mmHg	Pulse: 40-240 bpm
Resolution	±1 mmHg	±2 bpm
Accuracy	±3 mmHg	±1 bpm

Tabella 3.1: *Gima ABPM specifications*

Figura 3.3: *Gima ABPM*

3.1.2 Software

3.1.2.1 MATLAB

MATLAB stands for matrix laboratory. It is a multi-paradigm numerical computing environment and programming platform designed specifically for engineers and scientists. The heart of MATLAB is the MATLAB language, a proprietary programming language developed by MathWorks, that is a matrix-based language allowing the most natural expression of computational mathematics.

Figura 3.4: *Logo of Matlab*

3.1.2.2 Google Colab

Colaboratory, or "Colab" for short, is a product from Google Research. It is a free, hosted Jupyter notebook service that runs in the cloud. A Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Born in 2014 as an internal Google project, nowadays Colab allows students, data scientists or an AI researchers to write and

Figura 3.5: *Logo of Colab*

execute Python 2 and Python 3 in browser. It is especially well suited to machine learning, data analysis and education. It does not required any configuration and guarantees free access computing resources, including GPUs, and easy storing and sharing modalities: Colab notebooks are stored in Google Drive, can be loaded from GitHub and can be shared like Google Docs or Sheets by clicking a Share button. It works with most major browsers, and is most thoroughly tested with the latest versions of Chrome, Firefox and Safari.

In order to be able to offer computational resources for free, Colab needs to maintain the flexibility to adjust usage limits and hardware availability on the fly. Resources available in Colab vary over time to accommodate fluctuations in demand, as well as to accommodate overall growth and other factors. This means that overall usage limits as well as idle timeout periods, maximum virtual machine lifetime, GPU types available, and other factors vary over time. Colab does not publish these limits, in part because they can vary quickly. To satisfy users in need of faster GPUs, longer running notebooks and more memory, as well as usage limits that are higher and do not fluctuate as much, Google offers also a Pro version of Colab.

The GPUs available in Colab often include Nvidia K80s, T4s, P4s and P100s. There is no way to choose what type of GPU you can connect to in Colab at any given time. GPUs are sometimes prioritized for users who use Colab interactively rather than for long-running computations, or for users who have recently used less resources in Colab. As a result, users who use Colab for long-running computations, or users who have recently used more resources in Colab, are more likely to run into usage limits and have their access to GPUs temporarily restricted. Colab provide also the possibility to use Colab's UI with a local runtime running on their own hardware, in case of need.

Notebooks run by connecting to virtual machines that have maximum lifetimes that can be as much as 12 hours, but notebooks are also disconnect from VMs when left idle for too long. The amount of memory available in Colab virtual machines varies over time, even if is stable for the lifetime of the VM. Colab may sometimes automatically assigned a VM with extra memory to a user when Colab detects she is likely to need it.

3.1.2.3 Python and main libraries

Python is a modern programming language, with a simple and powerful syntax that facilitates its learning. This programming language has different and numerous application areas, such as the development of websites, the implementation of web and desktop applications, the creation of graphic interfaces, system administration, scientific and numerical calculation, databases, games, 3D graphics, etc. Developed by Guido Van Rossum, it is the heir of Unix language ABC. The main competitive advantages of Python are:

Figura 3.6: *Logo of Python*

- It is free. Python is completely free and can be used it and distributed without copyright restrictions. Despite being free, since over 25 years Python has a very active community, and constantly receives improvements that keep it updated.

- It is multi-paradigm. Python supports both procedural programming, which makes use of functions, and object-oriented programming. It also supports several elements of functional programming.

- It is portable. Python is a portable language developed in ANSI C. It can be used on various platforms such as: Unix, Linux, Windows, DOS, Macintosh, Real Time Systems, OS/2, Android and iOS mobile phones. This is possible because it is an interpreted language, so the same code can be executed on any platform as long as it has the Python interpreter installed.

- It is easy to use. Python is a high-level language that is both simple and powerful. The syntax and the various modules and functions that are already included in the language are consistent, intuitive, and easy to learn, and the design of the language is based on the principle of least astonishment (i.e. the "minor surprise": the behavior of the program coincides with what expected).

- It is full of libraries. Each installation of Python includes the standard library, that is a collection of over 200 modules to perform widely varied tasks, such as interaction with the operating system and the filesystem, or the management of different protocols. In addition, the Python Package Index allows you to download and install thousands of add-on modules created and maintained by the community.

- It is performing. Although Python is considered an interpreted language, programs are automatically compiled in a format called bytecode before they are executed. This format is more compact and efficient, and therefore guarantees high performance. In addition, various data structures, functions, and Python modules are implemented internally in C to be even more performing.

- It automatically manages memory. Python is a high-level language that adopts a garbage collection mechanism that automatically takes care of the allocation and release of memory. This allows the programmer to use variables freely, without having to worry about declaring them and allocating and releasing memory spaces manually (which is necessary in lower level languages such as C or C ++).

- It can be integrated with other languages. In addition to the classic interpreter written in C (and called CPython), there are also other interpreters that allow integration with several other languages. IronPython allows you to use Python within the .NET framework, to use its functions, and to interact with other .NET languages. In order to integrate Python and Java instead, you can use Jython. There are also other interpreters, such as PyPy: a highly performing implementation written in Python.

In order to develop this project, several Python libraries have been selected, imported and exploited according to the requirements and needs of the different tasks. The main ones are the following:

- NumPy. It is a library adding support for large, multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays.

- SciPy. It is a library used for scientific computing and technical computing. SciPy contains modules for optimization, linear algebra, integration, interpolation, special functions, FFT, signal and image processing, ODE solvers and other tasks common in science and engineering. SciPy builds on the NumPy array object.
- Scikit-Learn. It is a machine learning library that features various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means and DBSCAN, and is designed to interoperate with the Python numerical and scientific libraries NumPy and SciPy.
- Keras. It is a high-level neural networks API, written in Python and capable of running on top of TensorFlow, CNTK, or Theano. It was developed with a focus on enabling fast experimentation, by leveraging user friendliness, modularity and easy extensibility.
- TensorFlow. It is a free and open-source software library for dataflow and differentiable programming across a range of tasks. It is a symbolic math library, and is also used for machine learning applications such as neural networks.
- MatPlotLib. It is a plotting library for NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits.
- Seaborn. It is a data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.
- WandB. It is a library for the integration of Weights & Biases on several supported frameworks. It allows to design and keep track of machine learning experiences. Further explanations can be found in the following section.

3.1.2.4 Weights & Biases

Weights & Biases is a framework-agnostic experiment tracking platform for deep learning. Its tools make collaborative deep learning easier by organizing and analyzing machine learning experiments and notes in a shared workspace, tracking all the code and hyperparameters, and visualizing output metrics. Each time one run a script instrumented with wandb, it saves its hyperparameters and output metrics. It allows to visualize models over the course of training, in real-time, and compare versions of your models easily. It also automatically track the state of the code, system metrics, and configuration parameters.

Figura 3.7: *Logo of WandB*

3.2 Development

3.2.1 Dataset

There are two main publicly available databases containing both PPG and BP signals. The first one is the PPG-BP Database, that was recently published and it contains

PPG signals collected along with BP readings from patients admitted to the Guilin People's Hospital in Guilin, China. It includes data collected from 219 subjects, aged 21-86 years, with a median age of 58 years, covering several diseases including hypertension, diabetes, cerebral infarction, and insufficient brain blood supply. The second database is the third version of the Multiparameter Intelligent Monitoring in Intensive Care (MIMIC) Database, which contains thousands of recordings of multiple physiologic signals such as arterial blood pressure, PPG signals, ECG signals, respiration, and many other additional waveforms simultaneously collected. MIMIC, both in its second and third version, has been used to assess pulse transit times in several studies, mainly by processing either PPG-ECG couple of signals or PPG only. It is worth noting that though multi-parameter hospital monitors collect various physiological parameters "simultaneously", the electronic hardware and software filtering on the measured signals produce additional time delays up to 500 milliseconds. This delays can have a significant effect on the calculation of blood propagation time intervals. Therefore, it is not recommended to use the whole MIMIC database to calculate VTT, PTT or PAT, because signals are not perfectly synchronized. Furthermore, both PPG-BP Database and MIMIC III do not include phonocardiographic signals synchronized with photoplethysmographic ones.

The unavailability of dataset useful for this project led to the necessity to create from scratch a new database, containing perfectly synchronized PPG and PCG signals so that VTT could be reliably estimated. As previously disclosed, the first phase of the project comprised the training of the neural network to detect critical reference points in temporal series, while the second one the computation of VTT for the estimation of BP through a simple linear regression approach, the third one the actual implementation of the BP estimator. In general, convolutional neural network training require bigger databases with respect to regression models. In fact, while linear regression models must learn a simple relationship between the independent variable and the dependent ones by estimating a restrained number of weights, CNN face the same input-output mapping problem in a definitely more complex context. A neural network has several inputs feeding several layers and each layer contains several weights that must be suitably estimated in order to guarantee a proper learning. Usually, this aim can be satisfactorily accomplished by employing big datasets, containing many samples, as much variate and heterogeneous as possible. This variety has a crucial positive effect: it allows to prevent overfitting problem, that happens when the network learn too precisely the training dataset features, loosing the ability to generalize its knowledge. When overfitting occurs, the net performs well on the training data, but since the relationships it learned is not enough general, its performance is considerably worse on independent data that have not been used in the learning process. Although there was the necessity of collecting many recordings, the blood pressure measurements needed for the regression would have been collected by means of the cuff-based oscillometer that takes a long time to complete each sampling and can cause discomfort in the examined subjects. For these reason, was taken the decision to build two distinct datasets, one for each phase of the project and based on substantially different protocols. This choice allowed to gather a large volume of recordings for the CNN training by exploiting Soundi without excessively fatiguing the voluntary subjects.

3.2.1.1 Neural network database

Protocol The protocol for the data acquisition can be divided in two main phases: system setup phase and data acquisition phase.

In setup phase, the participant was asked to seat comfortably on a chair for about 5 minutes. As previously described, Soundi must be applied on the subject chest, in close contact with his skin, in correspondence of his heart. At setup time, the clinical operator was required to position the device and started it in setup mode. In this functional mode, the original signals were delivered through BLE to the host computer. Data collected in this phase were show in real-time on the screen and used by the operator to visualize and inspect qualitatively the signals, with the objective of detecting the eventual presence of noise, interference or malpositioning. In case of insufficient quality of the signals, the device was carefully repositioned, as long as the signals were arbitrary considered acceptable by the operator. This quality check phase led to sensible signal quality improvement, thus easing successive post-processing phase. Once the signals were considered to be satisfactory, the operator positioned in a definite way the device on the subject skin through a medical adhesive tape. Then, the device functioning mode was switched to normal acquisition and the starting time of the registration was set, as well as its duration.

The acquisition phase was executed with two different modalities, in order to increase the dataset variability. The first modality implied the subject at rest, while the second one contemplated data acquisition after the subject performed an intense physical activity. In at-rest acquisition phase, the participant was asked to seat comfortably for exactly 15 minutes, wearing the developed device, and to remain still and in silence, in order to limit the noise, especially on the acoustic signal. In after-activity data acquisition, the subject was required to run on a treadmill at 10 km/h for 10 minutes. Then, he was asked to sit in silence wearing the device for 15 minutes as well. These second modality allowed to introduce in the dataset recordings characterized by a higher heart rate (HR) and shorter cardiac cycles. Furthermore, since during the registration the HR was physiologically decreasing because of rest condition, a complete spectrum of HR and pulse wave lengths was recorded and therefore used for training the CNN.

Dataset description By following the above explained protocol, about 5 hours and 15 minutes of recording were collected, 30 minutes of which were generated with after-activity protocol modality. This dataset is then integrate with 1 hour and 30 minutes of recordings belonging to Porro's database [1] and generated in 2017 with the same at-rest protocol modality. Recordings were arbitrary selected according to the quality of the signals. The final database was constituted by about 5 hours and 10 minutes of signals and involved a total of 15 adult volunteers, of which 6 females and 9 males. The age of participants ranged 24-37 years old and no one presented known cardiovascular anomalies.

3.2.1.2 Regression database

Protocol As for the NN database, also for regression database the acquisition protocol can be divided in two main phases: again, the system setup phase and data

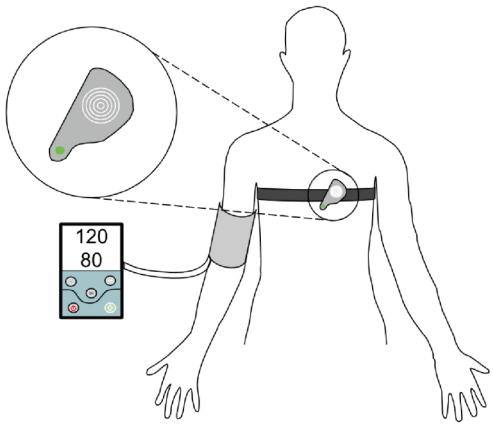

Figura 3.8: *Measurement protocol required Soundi to be applied on the subject chest, in close contact with the skin and in correspondence of the heart, and the subject to wear the GIMA ABPM oscillometer on the left arm.*

acquisition phase. The system setup phase was basically identical, except for the fact that the subject was required to wear the GIMA ABPM oscillometer on the left arm. One at-rest BP measurement was also performed in this phase and recorded to be used as a reference for successive measurements. The acquisition phase was instead profoundly different. Since to train a regression model predicting BP from VTT several couples of corresponding values are needed, two approaches could have been possibly adopted: 1) guarantee high inter-subject variability by recruiting many volunteers and testing them in normal conditions to attempt to create a very general model or 2) invest available time resources in increasing intra-subject variability, by inducing appreciable BP variation through some technique, to train a subject-specific predictor similar to analogous regression models using PAT and PTT found in literature. After a test of the first approach, explained in section 5, the second approach was chosen. Among common interventions proposed in the literature for perturbing BP (Table 2.3), a supervised physical exercise was employed, since it has been shown to cause a sensible increase in both SBP and DBP up to 40 mmHg.[3]. Subjects were asked to sit on a cycloergometer for 60 minutes, wearing both Soundi and the holter. They were asked to cycle for 3 minutes at constant speed of 15 km/h, then to stop for 2 minutes. At the beginning of the pause, a BP sampling, lasting about 30 seconds, was executed and the subject was told to remain stable with no movement and not to speak because the consequent vibrations in their chest could have corrupted the PCG signal. The remaining 1.5 minute of pause was introduced with a double purpose: to allow the subject to rest and to avoid arm congestion, that produce discomfort and potentially unreliable BP measurements. The described process was repeated across three phases of 10 minutes each: in the first phase the ergometer was set at minimum resistance level, in the second phase at intermediate resistance level, in the third phase at maximum resistance.

Dataset description The dataset was built by extracting the signal segments after the beginning of each pause, in a temporal window of 30 seconds. Each chunk was then paired with the corresponding BP measurement. One male subject of 25 years

old participated at the experiment. Thus the dataset was composed by 12 segments of 30 seconds each per signal type and 12 simultaneous numeric measurements.

3.2.2 Pre-processing

After the on-board acquisition and storage of the PCG and PPG signals, data were transferred on a PC and then processed offline with Matlab. In this phase Matlab had been chosen for several reasons:

- Matlab offers a complete development environment and several tools that allow easy handling of signals, such as visualization tools, tools for accessing intuitively to data and matrices, a comprehensive and user-friendly graphic user interface, an unbeatable and efficient library support.
- Despite Google Colaboratory is a powerful tool and do not overwhelm your CPU/GPU because it works on an online virtual machine, its run-time is excessively limited and when it expired the entire workspace is lost, if not preventively saved.

Therefore, since this phase did not require high computational efforts, Matlab was preferred.

After a careful qualitative inspection of the signals, they were filtered to clean the noise out. Infinite Impulse Response (IIR) filters were used, because of their greater computational efficiency than Finite Impulse Response (FIR) filters. IIR filters, however, have non-linear phase responses, which may lead to distortions in signals that surely will affect the accuracy of VTT measurements. To avoid this pitfall, IIR filters were applied bidirectionally, which induced a zero-phase response. Both PCG and PPG were filtered by using a band-pass Butterworth digital filter of 4th order with different cutoff frequencies according to the signal type. Optical signal was deprived of frequencies below 1Hz and above 8Hz, while acoustic one was filtered in a slightly more conservative way, attenuating components below 10 Hz and above 40 Hz. These filter configurations allowed to remove both the low frequency components due to breathing, and the high frequency effects of muscle vibrations and residual heart sounds. Such invasive filtering was executed in order to keep the waveform as smooth as possible with the objective to ease the development and the accuracy of an automatic labeling algorithm able to detect the reference points needed to compute the vascular transit time. Even though clinically useful content of PCG resides also at higher frequencies[46], keeping them was considered a complicating element, since they seems to do not significantly contribute to the energy envelope used in the labeling process explained in the following section.

3.2.3 Labeling

The objective of this first part of the project was, as previously explained, the identification of reference points on PCG and PPG signals that would have been used as a main ingredients for the computation of the vascular transit time (VTT), which would then have been used to estimate the BP by leveraging a regression-based predictor. To accomplish this task, a supervised learning approach was chosen. Supervised learning

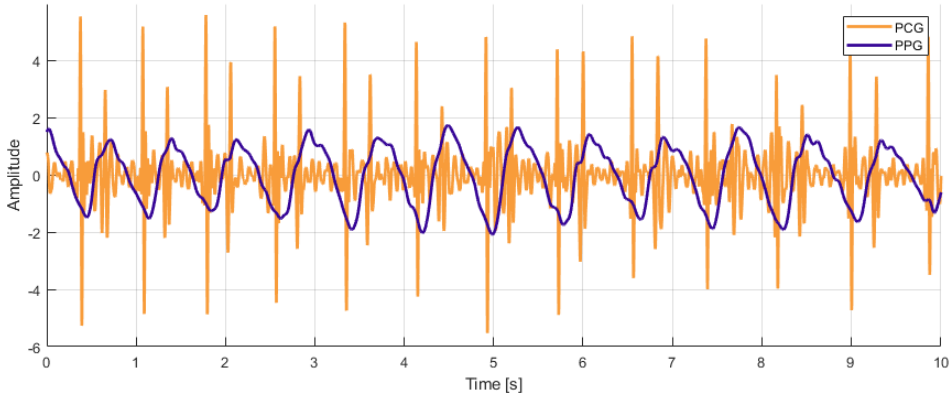

Figura 3.9: *Signals filtered through a zero-phase, band-pass Butterworth digital filter of 4th oreder. Cutoff frequencies were 1Hz and above 8Hz for PPG, 10Hz and above 40Hz for PCG.*

algorithms build a mathematical model of a set of data that contains both the inputs and the desired outputs, also known as "labels", that constitutes a supervisory signal. Learning of a prediction function is done through iterative optimization of an objective function measuring the error, thus the distance of the current predictions from the labels. Pulse transit time (PTT) and pulse arrival time (PAT), the two time-intervals analogous to VTT used to measure the delay of propagation of a pulse waveform, have been computed in literature by using three different PPG label types (Chapter 2), such as:[6]

- foot, that is the onset of the PPG peak, thus the minimum in the systolic phase;
- maximum positive slope point, that is the point with maximum gradient in the systolic ascending tract of the PPG;
- systolic peak, usually the maximum point of the signal in a cardiac cycle.

Usually, one of this reference points is arbitrary selected and compared to another reference point on a signal measured in a different source location, to determine the propagation delay. While PAT is computed as the temporal difference between a PPG reference point and the R peak of QRS complex in ECG, here the latter was substituted by S1 peak of PCG. Since VTT has not been widely explored, which is the best PPG reference point has not been studied yet, except in rare cases[3]. From this uncertainty born the necessity to investigate the reliability and the effectiveness of using each reference type. Thus the labeling process of the project aimed at identifying four different useful markers: three for PPG signal and one for PCG one. Since the whole database for neural network training included more than 20,000 cardiac cycles, a manual labeling, even though accurate, would have been tremendously time consuming. For this reason, a method for automatic labeling was implemented, hence yielding to a trade-off between labeling accuracy and invested time.

PPG foot Filtered optical signal had a simpler and smoother waveform with respect to acoustic signal. Due to this simplicity and easiness of interpretation, PPG was selected as the first signal to be labeled. In particular, PPG foot is trivial to detect,

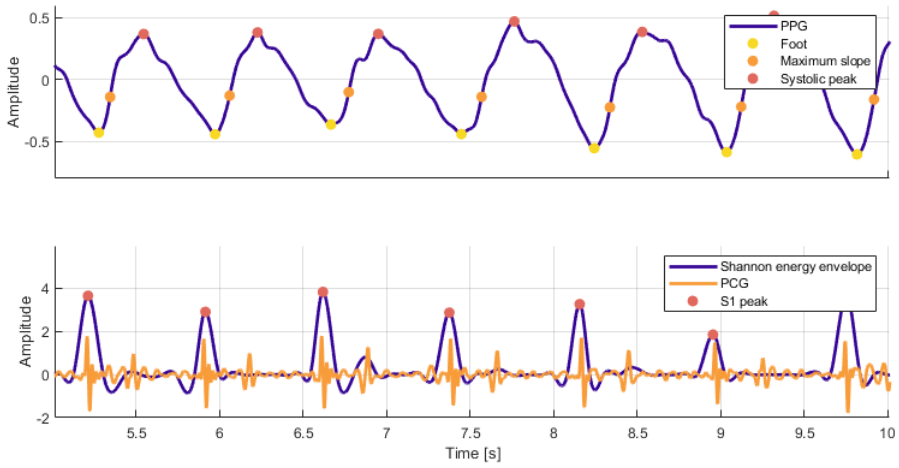

Figura 3.10: *Synchronous signals labeled. The labels of photoplethysmographic signal (PPG) were the foot, the maximum positive gradient and the systolic peak, corresponding to the minimum, the maximum of first order derivative and the maximum, respectively. The label of phonocardiographic signal (PCG) was the S1 peak, corresponding to the peak of the Shannon energy envelope.*

since, in general, it is just the minimum of each the PPG complex. First of all, signal was normalized by computing the z-score, ergo transforming the series to have null mean and unitary standard deviation. Then, the signal was further smoothed through a low-pass Gaussian filter, that removed residual, undesired noise spikes. After that, maximum peaks on the opposite signal were identified by accounting for local maximum prominence. The prominence of a peak measures how much the peak stands out due to its intrinsic height and its location relative to other peaks. Prominence threshold was set to keep reasonably high the false positive rate and the sensitivity with the aim of maximizing the number of recognized peaks. This voluntary bias was secondly corrected by the introduction of a mechanism of peak suppression that accounted for the normal duration of a cardiac cycle in a specific track. More in detail, the heart rate variability (HRV) was computed as the difference between the time instants of a peak and its precedent one. HRV sequence was then adjusted by applying a winsorization, thus substituting the outliers with a specified percentile of the data, and filtered with a moving average, obtaining a reliable low frequency trend of the cardiac period across time. Finally, the consistency of each peak with the corresponding cardiac period value was evaluated: if three consecutive peaks were excessively closed together with respect to the cardiac period, the median one was deleted. A further correction was applied, substituting each index with the nearest minimum index in its neighborhood in the not-smoothed signal. The final sequence of peak indices corresponded to the definitive foot labels.

PPG maximum slope Determined PPG foot indices were used as reference for locating the maximum slope in the systolic phase. Basically, for each index was defined a time window, corresponding to half cardiac period, where to look for the highest peak of the smoothed first derivative of PPG. The indices obtained were considered

the correct ones.

PPG peak Unsurprisingly, an approach similar to that adopted for identifying the maximum slopes was adopted to detect the PPG peaks. This time, anyway, the algorithm seemed to be less reliable, due to the fact that the systolic peak was not always evident, sometimes merging with the diastolic one.

PCG S1 sound S1 sounds were detected by exploiting the advantage of PCG synchronization with PPG signal. In concrete terms, each maximum slope index was used as a reference to define a time window, preceding the index itself, where to look for the first heart sound. In fact, it is known that S1 sound corresponds to the ejection of blood by ventricles (Chapter 2), therefore must precedes the transit of blood pulse in the chest skin. This expedient allowed to develop a computationally light, fast, reliable detection algorithm by leveraging intrinsic time relationship between the two signals, without incurring in false positives due to S1 sounds mistaken for S2 ones. In order to easily identify the heart sound in an automatic way, an envelope of the Shannon energy of PCG was computed, highlighting the wanted peak,[52] whose maximum value index was finally registered. One may argues that this approach, by totally relying on the PPG labels, inherits its failings and misidentifications, aggravating the accuracy. That is definitely true, but the context of the application must be considered. Since the intermediate objective of the project was to estimate VTT, a misidentification of one reference point on a signal, would have inevitably invalidated also the corresponding detection on the other one, even if correct. Hence, the algorithm was somehow forced to labeling consistency between the two signals, avoiding, moreover, to check it in a successive step.

3.2.4 Dataset definition

After the identification of the reference points, the database was modified. While the labels remained the same, each signal was substituted with a less filtered version of itself. PPG segments were substitute with their original raw version, while PCG segments were replaced with a corresponding version filtered with a Butterworth high-pass filter of 4th order with cutoff frequency of 10 Hz. This action is due to the will to keep the signals as much similar as possible to the original ones, by preserving both the noble waveform and the noise, without sacrificing their typical shape. This precaution was meant to have a double effect:

- When a neural network learns, highly noisy signal associated to correct labeling can taught to NN to ignore noise, thus making it more robust.
- Less filtering implies a lower computational cost and less time resources spent for processing, hence an increased possibility to work in real-time with raw data.
- neural networks can find patterns in what we, as humans, would have commonly considered noise that can be useful to accomplish their task.

Anyway, while raw PPG was used, PCG was employed in a filtered form. The difference between the two behaviors can be easily explained: while the PPG waveform conserved

its typical profile, the PCG was almost unrecognizable, therefore the decision to clean it of low frequencies was taken.

3.2.4.1 Data structure

Once the database was ready, it was reorganized in a new dataset. Both the signals and the labels underwent a rearrangement.

Inputs Each signal segment was further split in smaller chunks, each 600 samples long. Since sampling frequency was 400 Hz, each chunk corresponded to 1.5 seconds of recording. The choice of the length was opportunely calibrated, as a compromise accounting for several factors:

- The chunks, that would have constituted the future training input samples of the NN, should have been as much as possible to improve the generalization ability and the learning results of the NN. As the overall duration of the recordings was fixed, numerous samples should have been inevitably small and simple samples.
- Since the dimension of the input of a convolutional NN influences the overall architecture complexity as well as its computational cost and its prediction speed, samples should have been short enough to guarantee a light and fast neural network able to work also in real-time.
- In this specific case, NN would have learned to identify a reference point by using a pattern extracted from the context. It would have been crucial to select a proper length for samples such that they would have included also this rich information.
- Chunks should have been long enough to include at least on complete cardiac cycle when the heart rate would have had the lowest possible value. 1.5 second samples allow indeed the inclusion of a whole cardiac period at 40 bpm, arbitrary chosen as lower limit for HR.

Targets Target structure, built by a rearrangement of the labeling, is strictly related to the network architecture and its functioning. Each 600 long sample was virtually divided in five windows, 120 each, thus 0.3 seconds. In each of these windows, at most one reference point per type could have been present, with a specific location inside the window. The idea was to rearrange the labeling such that in each window the new labeling would have been represented so that it expressed both the presence of a specific point and where it was situated. This aim was achieved by organizing the labels of each sample as a $W \times P$, where W is the number of windows, thus is equal to 5, and P is equal to 2 and it is composed by a binary classification bit (ID), hence expressing the presence or the absence of a specific landmark, and a value expressing the label position in the window with respect to the length of the window itself (LOC). Therefore, ID assumed value 0 when the label was not present in a specific window, 1 when it was. LOC instead, since it was not binary, could have assumed every value from 0 to 1, where 0 would have signified it was at the beginning of the window, 1 at the end of it and any intermediate value that it was at $LOC \times$

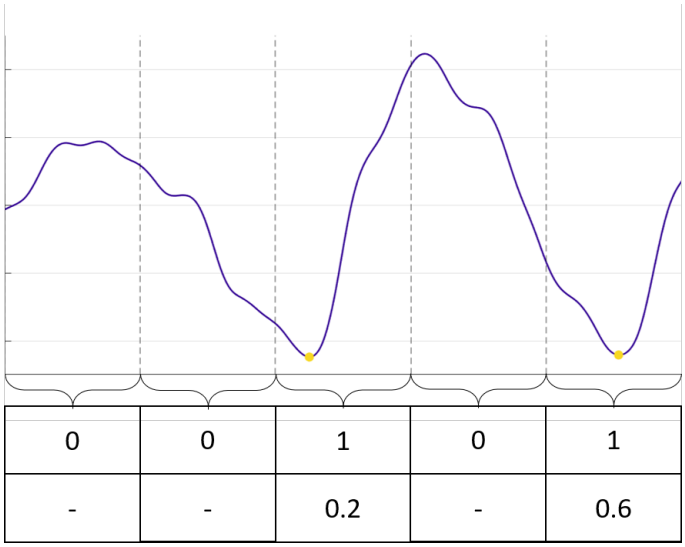

Figura 3.11: *Output example. Each sample of 1,5 seconds was virtually divided in five equidimensional windows, 0.3 seconds each. If an object falls into a window, that window is responsible for detecting that object. Each window is associate to two output neurons for each label type: one expressing a confidence scores for this identification (upper row), the other expressing the normalized position of the label inside the segment (lower row). The confidence score reflects how confident the model is that the segment contains an object. If no object exists in that segment, the confidence scores should be zero, otherwise the confidence score should be one. The normalized position instead could assume every value from 0 to 1, where 0 means the label is at the beginning of the window, 1 at the end of it and any intermediate value LOC that it is at* $LOC \times (Window\ Length)$ *with respect to the beginning of the window.*

$(Window\ Length)$ with respect to the beginning of the window, that corresponds to $[LOC + (Window\ Number)] \times (Window\ Length)$ with respect to the beginning of the sample, where $(Window\ Number) \times (Window\ Length)$ constitutes an offset. Here, the number of windows was chosen to have as few network outputs as possible and to avoid having more than one label per type in the same window. 0.3 second windows allow indeed this favorable condition as long as the heart rate of the subject is below 300 bpm, threshold that provides a generous margin of safety.

3.2.5 Part I: the Convolutional Neural Network

3.2.5.1 Dataset splitting

The whole neural network dataset contained about 22,000 samples, 600 units long each. It was then divided in three subsets, namely the training set, the validation set and the test set. The training set contained the data used to fit the deep learning model; the validation set included the data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyperparameters; the test set was composed by data used to provide an unbiased evaluation of a final model fit on the training dataset. Test and validation sets were defined by extracting 1,000 samples each, while the remaining 20,000 were dedicated to the training set. Every sample used to build a subset was not originally part of the same signal from where samples of another subset

were extracted from, thus independence among the subsets was guaranteed, as well as unbiased estimation of the performance.

3.2.5.2 Data augmentation and sample standardization

Regularization is a set of techniques aimed at reducing model overfitting. In this ensemble is included any modification made to a learning algorithm intended to reduce its generalization error but not its training error,[53] by adding prior knowledge to a model and decreasing the model variance. Data Augmentation is a regularization technique that approaches overfitting from the root of the problem, the training dataset. [53, 54] The performance of deep learning neural networks often improves with the amount of data available. Data augmentation is a technique to artificially create new training data from existing training data. This is done by applying domain-specific techniques to examples from the training data that create new and different training examples such that their label is preserved. The intent is to expand the training dataset with new, plausible examples. This means, variations of the training set images that are likely to be seen by the model. As such, it is clear that the choice of the specific data augmentation techniques used for a training dataset must be chosen carefully and within the context of the training dataset and knowledge of the problem domain. In one dimensional signals this encompasses augmentations such as additive noise, signal shifting, random erasing, etc. Data augmentation is typically only applied to the training dataset, and not to the validation or test dataset.

In this project the training set was augmented in order to improve the generalization ability of the NN and to make it more robust. In particular, the training set was doubled by adding noisy versions of the original samples. Samples to be augmented were selected with Bootstrap method, an independent sampling technique that requires replacement of the extractions. Each extracted sample was then randomly affected by a noise type among three:

- uniform noise with random amplitude;
- Gaussian noise with zero mean and random standard deviation;
- periodic noise, created by adding a random number of sinusoids with frequency in range 0.1-200 Hz and random phase.

After that, all the 40,000 training samples were standardized to have zero mean and unit variance. This step is particularly important, because it eased and boosted the NN learning. The reason why this happened is due to the fact that, after standardization, the network would not have had to learn also the signal mean to take it into account during predictions.

From now on every operation was executed on Colab and every script written in Python. The main reason of this choice was the need for greater computational power. Furthermore, Colab allows you to run multiple scripts simultaneously, parallelizing their execution and saving a lot of time.

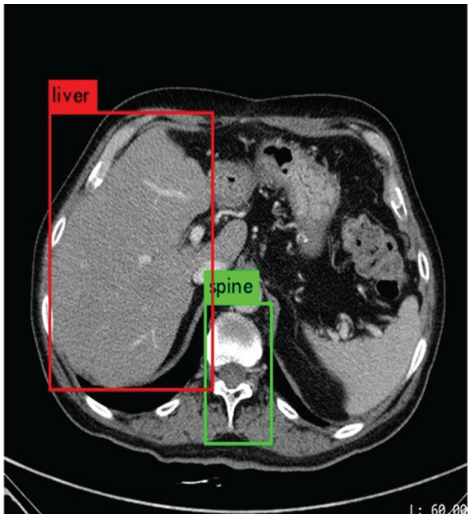

Figura 3.12: *Example of YOLO application in medicine.[11]*

3.2.5.3 The problem as a 1D object detection task

Object detection is a computer vision technique for locating instances of objects by usually leveraging deep learning. Thus, object detection comprehends both a classification task and a localization task. While object detection is usually applied to two dimensional frames, such as images or videos, in this project the concept has been extended (or shrunk) to fit a mono-dimensional context. In this landscape, the "frame" is constituted by the 1,5 second samples, while the objects to be detected are the different reference points of PPG and PCG. The two elemental tasks of object detection were merged here in a unique neural network able in one shot to both identify the presence of a landmark and locate it in the space.

3.2.5.4 Architecture design hypothesis

The designed neural network receives as input samples in format $1 \times 600 \times C_{\text{input}}$, where C_{input} is a variable number of channels according to the application, which will be explained later. The output instead has the format of the targets, thus $5 \times 2 \times C_{\text{output}}$, where C_{output} is a variable number of target classes, which will be explained later, too.

Two are the main architecture hypothesis that were evaluated, inspired by YOLOv2 algorithm and its lighter version, Tiny YOLOv2.[9] YOLO is an acronym for "You Only Look Once"[55], a concise expression that underlines its ability to perform 2D object detection phases, thus both classification and localization, in a single step, hence by looking only one time at the image frame. It is accurate and extremely fast, so much that its third version YOLOv3[56] currently represents the state-of-the-art for real-time object detection in bi-dimensional contexts. One YOLOv2 implementation, reported in the paper "YOLO9000: Better, Faster, Stronger" in 2016, relies on a custom architecture called Darknet-19. Similarly to the VGG models, it uses mostly 3×3 filters and double the number of channels after every pooling step. It uses also global average pooling to make predictions as well as 1×1 filters to compress the feature representation between 3×3 convolutions. It includes batch normalization to stabilize training, speed up convergence, and regularize the model. It is formed by

Type	Filters	Size/Stride	Output
Convolutional	32	3 × 3	224 × 224
Maxpool		2 × 2/2	112 × 112
Convolutional	64	3 × 3	112 × 112
Maxpool		2 × 2/2	56 × 56
Convolutional	128	3 × 3	56 × 56
Convolutional	64	1 × 1	56 × 56
Convolutional	128	3 × 3	56 × 56
Maxpool		2 × 2/2	28 × 28
Convolutional	256	3 × 3	28 × 28
Convolutional	128	1 × 1	28 × 28
Convolutional	256	3 × 3	28 × 28
Maxpool		2 × 2/2	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Convolutional	256	1 × 1	14 × 14
Convolutional	512	3 × 3	14 × 14
Maxpool		2 × 2/2	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	512	1 × 1	7 × 7
Convolutional	1024	3 × 3	7 × 7
Convolutional	1000	1 × 1	7 × 7
Avgpool		Global	1000
Softmax			

Figura 3.13: *DarkNet19 classification architecture. YOLO authors trained the DarkNet19 for object detection in two steps. First of all, they trained the architecture in figure for object classification. This network had 19 convolutional layers and 5 maxpooling layers. Then they modified this network for detection by removing the last convolutional layer and instead adding on three* 3×3 *convolutional layers with 1024 filters each followed by a final* 1×1 *convolutional layer with the number of outputs needed for detection.*

19 convolutional layers and 5 maxpooling layers. Tiny YOLOv2 is somehow instead inspired by LeNet5[57], one of the earliest convolutional neural networks and promoted the development of deep learning. Developed in the '80s and reported in a paper in 1989, LeNet5 was basically an alternate sequence of convolutional and pooling layers for image recognition.

Starting from these two YOLO implementations, many network structures were designed, analyzed and tested. Despite this variety of proposals, all the architectures had some constant characteristics, taken as assumptions to limit the dimension of the hypothesis space to explore:

- Convolution layers had kernels of size either 3 or 1. Although in literature 5 and 7 dimensional kernels are widely used for complex tasks, here the kernel

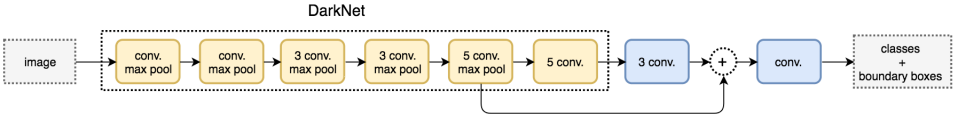

Figura 3.14: *DarkNet19 final architecture.*

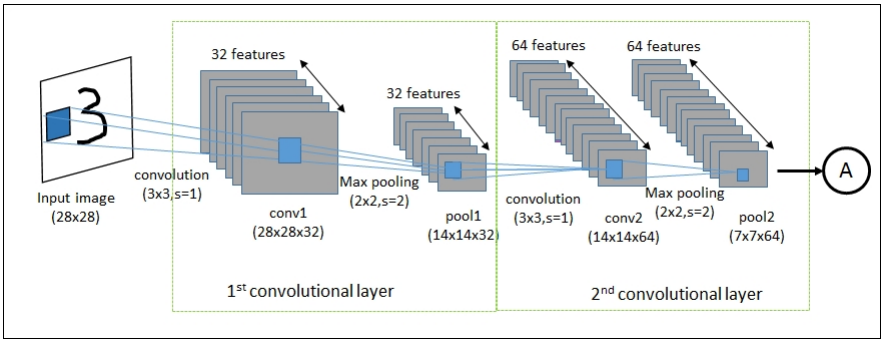

Figura 3.15: *LeNet5 architecture. It is an alternate sequence of convolutional and pooling layers for image recognition.*

size was kept small to induce the first layers to learn very simple patterns. This expedient allowed superficial layers to avoid learning high level features that could potentially have led to easy overfitting. This regularization choice allowed to manage the learning ability of the network by simply varying the total number of layers. The deeper the layers, the more specialized they were in detecting complex features. While 3-sized filters had the specific aim to extract patterns from data through convolution, the 1-sized layers had a radically different scope. Alternating them to 3-sized ones reduced the features space from preceding layers. Here 1×1 convolution was used to spatially combine features across feature maps generated by previous convolutions. The net effect was the forced selection of a lighter, less prone to overfitting, feature space as well as a faster and more computational efficient architecture.

- Convolution layers had unitary stride length in both directions. Since the objective of the network was to identify small, very specific targets, the stride length was fixed to one to avoid loss of precious information.

- Convolution layers had a progressively higher number of filters that had the purpose of extracting patterns through combination of an increasing number of feature. Indeed, the bigger the feature space, the higher the number of possible combination leading to high-level features.

- Convolution layers had a padding function that allowed an application of filters along all the input sample length that would otherwise have increase the information loss.

- Convolution layers had a leaky rectified linear activation function (Leaky ReLU). Literature has shown that this choice increase learning speed with respect to simple rectified linear activation.

- Max Pooling layers are used for downsampling, to extract locally relevant features and forget the others, simplifying the future space. This was done in part to help over-fitting by providing an abstracted form of the representation. As well, it reduced the computational cost by reducing the number of parameters to learn.

- Final dense layer had a linear activation function, since it would have to output a continuous numerical value, at least for the location, even though bounded between 0 and 1.

So, in summary, the template of the neural network structure can be generalized as follows:

1. Input layer of shape $1 \times 600 \times C_{\text{input}}$.
2. Several hidden layer blocks composed by a combination of a variable number of convolutional layers with a Leaky ReLU activation function followed by a maxpooling layers.
3. Dense output layer of shape $5 \times 2 \times C_{\text{output}}$ with linear activation function.

3.2.5.5 Chain of networks

As for the number of channels C_{input} and C_{output}, all the designed neural networks were able both to receive as input a combination of synchronized time series and to output more than one reference point type at time. In fact, the ability of the designed neural networks to predict simultaneously more than one label on the same signal was tested. In particular, the networks working with PPG were used to detect in parallel peaks, maximum slopes and feet. In this occasion, C_{output} was equal to 3, while, for NN predicting only S1 sounds of PCG, C_{output} was equal to 1. Furthermore, for choice the networks were made to predict the labels on just one signal type at time. This means that once a prediction was done for one signal type, this prediction could have been informative and useful also to process the other signal type, similarly to what had been done previously for automatic signal labeling. For example, once a NN had labeled a PPG chunk by detecting its peaks, this labeling could have been used to provide useful information to the NN that successively would have labeled the corresponding PCG chunk by detecting its S1 sounds. In this context, neural networks working on different signals would have been organized in chain. This chain was actually implemented and tested, in parallel with non-chained, independent structures for each signal type. In order to feed the second NN of the chain, the true labeling of the previous signal was converted from the format $5 \times 2 \times C_{\text{output}}$ to a binary signal with format $1 \times 600 \times C_{\text{output}}$. Each binary channel element was equal to 1 when label $\text{C}_{\text{output,i}}$ was present, 0 otherwise.

3.2.5.6 Training, testing and tuning

As anticipated, several architectures had been trained and then several parameters had been tuned during the project. As for the architecture, in order to keep the writing as concise as possible, a coding schema will be now introduced. From now on, each architecture will be represented by a sequence of numbers of variable length. While the length represents the number of inner blocks constituting the network, the values represent the number of convolutional layer in each block. Each block is implicitly followed by a maxpooling layer. The following structures had been tested:

- 11

- 113
- 1133
- 11335
- 113355
- 1111113

Once the most promising neural network architecture was identified, several parameters were tuned:

- Overall number of filters in convolutional layers.
- Number of filters in 1D convolutional layers.
- Maxpooling filter dimension, either 2 or 3.
- Learning rate, either 0.001 or 0.01 or 0.1.
- Batch size for training process, either 32 or 64 or 128.
- Number of input channels, either 1 or 2 or 3 for PCG networks.
- Number of output channels, either 1 or 3 for PPG networks.

The training process was performed on the training set, while the overfitting trend was assessed by comparing the performance of the net on the training set with the one on the validation set, through a specifically designed loss function that will be explained in the next section. Each model was trained with an indefinite number of epochs, upper bounded to 1,000. In fact, the learning process were early stopped once the network performance on the validation set was getting worse. Since the learning process was not very smooth and stable, gradient local minima could have stopped the training too early. To avoid this condition, a patience mechanism was implemented, that would have allowed training stop just if at least five epochs without validation loss improvements had occurred. Training was performed by using Adam optimizer, an adaptive learning rate optimization algorithm that has been designed specifically for training deep neural networks. [58] Furthermore, each training step of every experiment was recorded and logged to Weights and Biases platform for visualization and comparison with other tested architecture.

3.2.5.7 Evaluation functions

Loss functions express the discrepancy between the predictions of the model being trained and the actual problem instances. They are used in training phase by back-propagation algorithm to adjust and fine-tune the weights of a neural network, proportionally to the prediction error. Evaluation metrics instead are other functions, used as well to assess and judge the performance of a model, that are not used for optimization and learning. These evaluation functions can be broadly classified into two major categories depending upon the type of learning task one is dealing with: regression

losses and classification losses. As previously explained YOLO-like networks perform both an identification task and a localization task. Here the former was treated as a binary classification problem, while the latter as a regression problem. The complexity of this approach required a suited, custom loss function, able to take both aspects into consideration. Therefore the loss function was defined as the sum of an identification loss with a localization loss:

$$Custom\ Loss = Identification\ Loss + Localization\ Loss$$

$$Identification\ Loss = \frac{1}{N}\sum_{i=1}^{N} |y_{i,true}^{id} - y_{i,pred}^{id}|$$

$$Localization\ Loss = \frac{1}{N_{TP}}\sum_{i=1}^{N} \left(T_{id}(y_{i,pred}^{id}) \times y_{i,true}^{id} \times |y_{i,true}^{loc} - y_{i,pred}^{loc}|\right)$$

$$T_{id}(a) = \begin{cases} 1, & \text{if } a \geq 0.5 \\ 0, & \text{if } a < 0.5 \end{cases}$$

Here, the identification loss is the mean absolute distance between the prediction of the presence of a reference point and the corresponding true value, ergo the mean absolute error (MAE). In this case, the prediction made by the NN measures how confident it is of the presence of a specific landmark. The localization loss instead measures the prediction error of the point location, thus should be computed only when the predicted point is actually present in a specific window. For this reason it was defined as the MAE of localization when a predicted reference point was present, but shrunk towards zero when absent. N_{TP} is the total number of true positives.

To evaluate separately regression and classification problems, two kind of metrics were defined. Identification error was measured as the complementary of accuracy of point presence prediction. Prediction outcome was considered positive if the confidence was above a threshold, hence replaced with 1, else replaced with 0.

$$Identification\ Error = \frac{1}{N}\sum_{i=1}^{N} |y_{i,true}^{id} - T_{id}(y_{i,pred}^{id})|$$

The distance between a label location prediction and its true position was considered an error when above a certain threshold measured in milliseconds. During the training process the NN performance was tracked by using several tolerance values.

$$Localization\ Error = \frac{1}{N_{TP}}\sum_{i=1}^{N} \left(T_{id}(y_{i,pred}^{id}) \times y_{i,true}^{id} \times T_{loc}(|y_{i,true}^{loc} - y_{i,pred}^{loc}|)\right)$$

$$T_{loc}(a) = \begin{cases} 1, & \text{if } a \geq tolerance \\ 0, & \text{if } a < tolerance \end{cases}$$

3.2.5.8 Noise sensitivity test

Once the training process was concluded, the tuned models that performed better were stored. These models were then used to reach the target of the second part of the project, that was the extraction of features by using the detected points. During this phase, the test set was employed. In order to test the robustness of the model, a noise sensitivity test was performed. Performances of models were evaluated several times, with an increasing level of signal to noise ratio (SNR), where SNR was computed as:

$$SNR(dB) = 10 \log_{10} \frac{P_{signal}}{P_{noise}}$$

$$P_{signal} = \mathbb{E}[\, signal^2 \,]$$

$$P_{noise} = \sigma^2_{noise}$$

$$noise \sim \mathcal{N}(\, 0, \, \sigma^2_{noise})$$

Where P_i represents the power of i and $\mathcal{N}$ is a Gaussian distribution.

3.2.6 Part II: feature extraction from predictions

3.2.6.1 VTT extraction

Vascular transit time was estimated as the distance between S1 sound and the corresponding reference point of PPG. Since the best PPG reference point for VTT computation and BP estimation has not been studied yet, VTT was computed in three different ways, according to the PPG reference point used, to investigate these relationships. First of all, samples were used as input for both the PPG NN and the PCG NN. Then, per each sample, VTT_i was computed as the difference between t_i, where i is a PPG reference type, and the nearest $t_{S1} < t_i$. After that, VTT outliers were removed by using very generous thresholds, specifically defined for each PPG reference type.

3.2.6.2 HR extraction

Also the heart rate was similarly extracted: per each reference type of each sample, HR_i was computed as the difference between t_j and the nearest $\widetilde{t_j} < t_j$. After that, HR outliers were removed by using a threshold-based method as well.

3.2.7 Part III: BP estimation through VTT

3.2.7.1 Dataset splitting and extraction of parameters

Best performing models were then used also to reach the target of the third part of the project, that was the estimation of blood pressure by leveraging the computed features through a regression approach.

The regression dataset contained data from one subject and it was composed by 12 signal records of 30 seconds each. Each track was divided, as previously seen, in segments of 1.5 seconds each, corresponding to 600 units, in order to feed the best

performing neural network models. Both VTT and HR were extracted and taken into consideration in the regression models. VTT was computed similarly to how was explained before, but not exactly in the same way. As previously, at the beginning samples were used as input for both the PCG NN and the PCG NN, but, after this step, a correction to eliminate a particular type of error was applied. In fact, since each track was divided in smaller segments, it could have happened that some PPG peaks were divided in two distinct segments. In general, this was not a problem, but when the separate tract was the systolic phase, it could happen that two predictions for the same reference point were made in two contiguous segments. For this reason was introduced a non-maximum suppression mechanism able to solve the problem and improve the prediction abilities. This mechanism was designed to identify the presence of more than one reference point for each type in a window of about 400 ms, far less than a cardiac cycle duration. Among the identified, only the one with the highest confidence score was kept and considered as prediction for the successive steps. Then, per each sample, $\mathrm{VTT_i}$ was computed as the difference between $\mathrm{t_{S1}}$ and the nearest $\mathrm{t_i}$. Finally, VTT outliers were removed by using the same threshold-based method as before. Similarly, instantaneous heart rate was computed as difference between two contiguous S1 peaks divided by the sampling frequency. Both VTT and HR was then averaged over the entire track. The final result was a subject-specific sequences of 12 feature sets, each including three VTT computations and one HR value, associated to an equal number of corresponding blood pressure set, each including systolic and diastolic blood pressure values.

3.2.7.2 The regression models

In order to understand which configuration of parameters worked best for assessing blood pressure values, several models were tested. Given $\mathrm{l} \in$ (SBP, DBP) and $\mathrm{i} \in$ (foot, max slope, peak), the evaluated models were the followings:

- Simple linear regression from each $\mathrm{VTT_i}$ to $\mathrm{BP_l}$

$$BP_l = a_i \times \phi_j(VTT_i) + b$$

- Simple linear regression from HR to $\mathrm{BP_l}$

$$BP_l = a_i \times HR + b$$

- Multiple linear regression from $\mathrm{VTT_{foot}}$, $\mathrm{VTT_{max\ slope}}$, $\mathrm{VTT_{peak}}$ to $\mathrm{BP_l}$

$$BP_l = \sum_{i \in (foot, maxslope, peak)} a_i \times \phi_j(VTT_i) + b$$

- Multiple linear regression from $\mathrm{VTT_{foot}}$, $\mathrm{VTT_{max\ slope}}$, $\mathrm{VTT_{peak}}$, HR to $\mathrm{BP_l}$

$$BP_l = \sum_{i \in (foot, maxslope, peak)} a_i \times \phi_j(VTT_i) + HR + b$$

- Multiple linear regression from the principal components obtained by performing a PC analysis on the parameters, PC($\mathrm{VTT_{foot}}$, $\mathrm{VTT_{max\ slope}}$, $\mathrm{VTT_{peak}}$, HR), to $\mathrm{BP_l}$

$$BP_l = \sum_{i=1}^{N_{PC}} a_i \times PC_i(\phi_j(VTT_{\text{foot}}), \phi_j(VTT_{\text{max slope}}), \phi_j(VTT_{\text{peak}}), HR) + b$$

  where $\phi_j(x)$ is extracted from a set of functions designed to created a non-linear feature space:

$$\phi(x) = \begin{cases} x & identity \\ \frac{1}{x} & reciprocal \\ \log x & logarithm \end{cases}$$

Furthermore, tests were done also by substituting, for each sample, $\mathrm{VTT_i}$ in the formulas with the average of the three $\mathrm{VTT_i}$, computed with the three labels with respect to S1 peak.

All the models were assessed through leave-one-out cross-validation approach to make the most of the limited amount of available BP samples. This expedient allowed to train each model on each of the 12 samples, by first calibrating the regression parameters on 11 samples, then predicting the remaining one. This process was repeated 12 times for each model. The predictions of each model were evaluated computing mean absolute error, mean squared error and coefficient of determination with respect to the reference measurements collected through the medical, cuff-based, pressure device.

Capitolo 4

Results

4.1 Part I: the Convolutional Neural Network

In the following are explained the creative process and the working pipeline followed during this project to design and implement the neural network that accomplish the main objectives. Together with the description of the steps done, also statistics and insights are reported.

4.1.1 PPG Net: training results

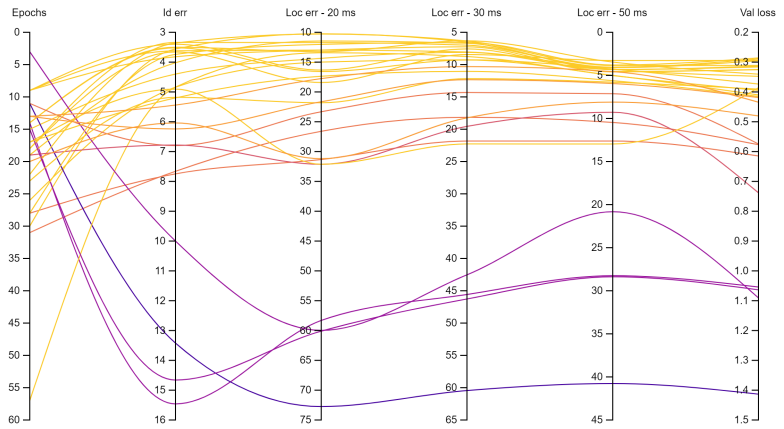

Figura 4.1: *Results of some architecture experiments done for object detection in PPG signals. "Epochs" is the number of training iterations, "Id err" is the identification error, the complementary of unbalanced accuracy. "Loc err" is the number of malpositions above a certain threshold expressed in milliseconds, it is a measure of dispersion of estimates around targets. "Val loss" is the value of the optimization function for the validation set.*

In figure 4.1 are graphically represented the results of some architecture experiments done for segmenting PPG signals. In the following paragraphs, some tracks are isolated and analyzed to give a complete overview of the results. The adopted approach for determining the best net structure consisted in a process of trial and error guided by prior knowledge derived from theoretical notions and intuition. Several structural

elements and parameters were tested in sequence, giving way to the ones able to explain the highest variance in the results and able to affect the most the network performance, such as the main architecture. Then, the resulting scaffolding was tuned by adjusting secondary parameters.

4.1.1.1 Main architecture

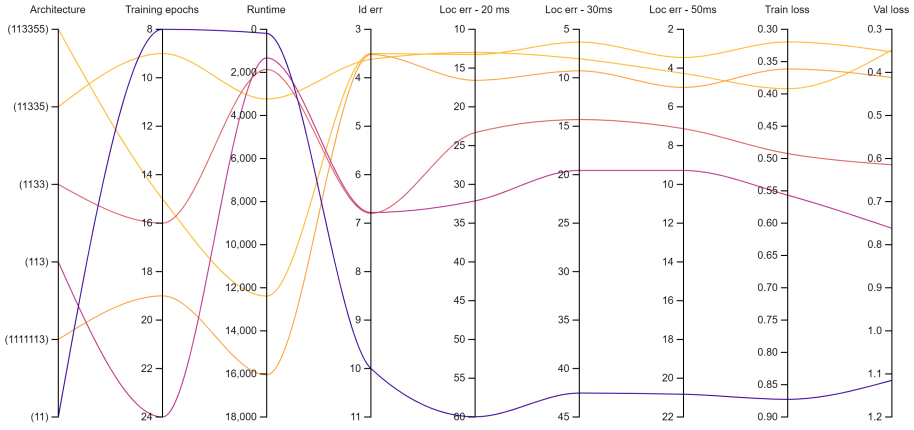

Figura 4.2: *Results of some architecture experiments done for object detection in PPG signals. Each architecture is represented by a vector of numbers of variable length. While the length represents the number of inner blocks constituting the network, the values represent the number of convolutional layer in each block. Each block is implicitly followed by a maxpooling layer.*

In figure 4.2 are reported the results of tested architectures trained to identify PPG foot, although relatively similar results were obtained also for the other labels, the maximum ascending gradient and the peak. Each architecture is represented by a vector of numbers of variable length. While the length represents the number of inner blocks constituting the network, the values represent the number of convolutional layer in each block. Each block is implicitly followed by a maxpooling layer. It is worth noting that, while simpler models struggled to reach good performances, deeper models, like 11335 and 113355, that were inspired by DarkNet architecture, and 1111113, aka Tiny YOLO, achieved promising results. This denotes a great complexity of the task. In this framework, the selected architecture had been the 11335 for two main reasons:

- It performed well on all the evaluation functions.
- It required few epochs to train and a very short runtime.
- It was relatively simple, thus computationally-undemanding.

4.1.1.2 Tuning

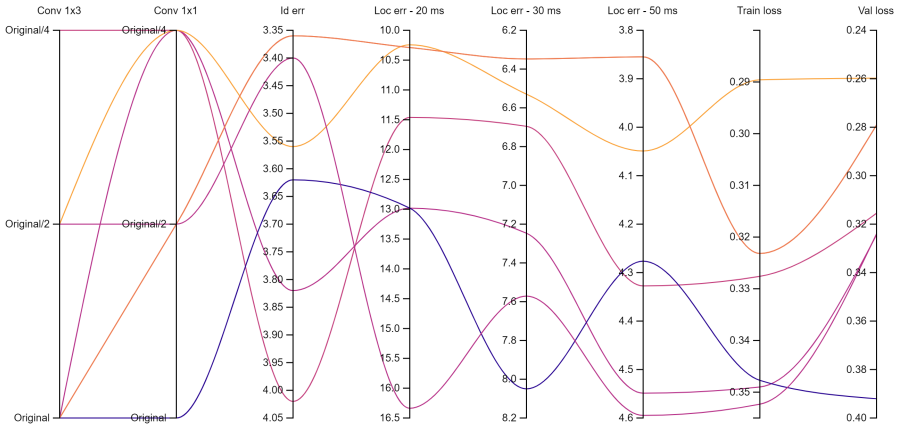

Figura 4.3: *Results of some architecture experiments done for object detection in PPG signals. "Conv* 1×1*" is the number of 1-sized convolutional layers with respect to the original DarkNet19. "Conv* 1×3*" is the number of 3-sized convolutional layers with respect to the original DarkNet19.*

After the selection of the network, the first parameter to be tuned was the number of filters in convolutional layers. The number of filters in the original DarkNet architecture was optimized to detect object in RGB images; here the framework is more simple, since we are handling monodimensional series, but not trivial as we are processing noisy signals. For this reason, architectures with a lower and reasonable number of filters were tested. Several combinations were evaluated, independently varying the number of kernels of size 3 and the ones with size 1. As shown in figure 4.3, with respect to the original architecture, the best configuration consisted in halving the number of 3-sized filters and dividing by four the number of 1-sized ones. This combination allowed to exploit anyhow a great quantity of filters for high-level feature extraction and to simultaneously reduce the feature space.

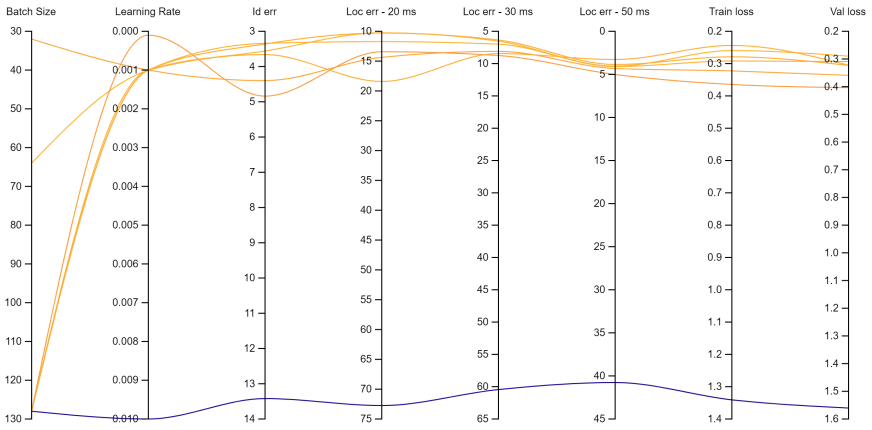

Figura 4.4: *Results of some architecture experiments done for object detection in PPG signals. "Batch size" is the dimension of the batches of samples used fore training. "Learning rate" is the weight given to the prediction error during back-propagation.*

Also the batch size and the order of the learning rate were tuned. The best combination consisted in a small learning rate, 10^-2, and a batch size of 128. Although not reported, also maxpooling filter dimension was optimize, resulting in 1-by-2 filters. Maxpooling presence is crucial to avoid both overfitting and huge number of filters.

4.1.1.3 Output Channels

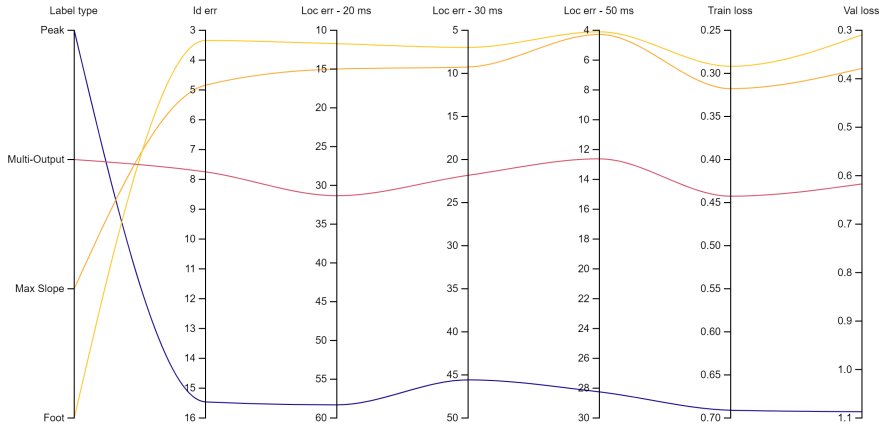

Figura 4.5: *Results of some architecture experiments done for object detection in PPG signals. "Label type" indicates the type of label the network has been trained for. Multi-output case includes all the labels.*

Finally, in figure 4.5 are reported the evaluation metrics for the current network trained with each output label type and with the combination of them. The best results on the validation set were obtained for the network optimized to recognize PPG feet, while less good results were obtained for PPG maximum ascending gradient and PPG peak. The main hypothesis for these different behaviors will be analyzed later. It is worth noting the performance of the multi-output network. Although its overall performance is worse than the one optimize to detect the PPG onset, its behavior on each single label type was accurately tested. It turned out that result of multi-output net were comparable to the corresponding single-output ones. For this reason the more complex NN was selected to be widely evaluated and exploited in the successive parts of the project.

4.1.2 PCG Net: training results

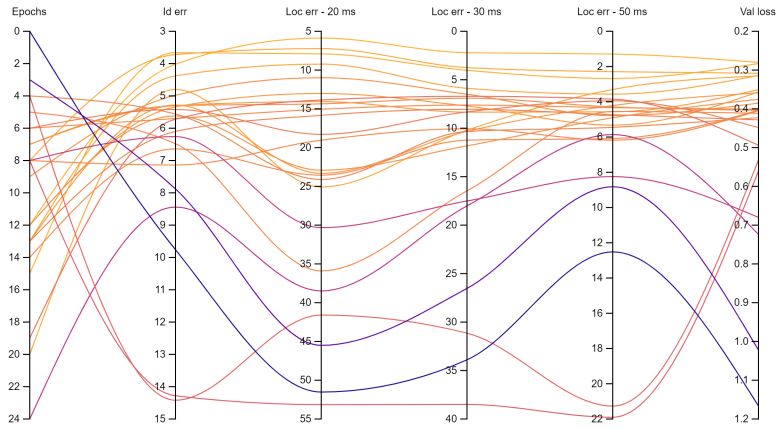

Figura 4.6: *Results of some architecture experiments done for object detection in PCG signals. "Epochs" is the number of training iterations, "Id err" is the identification error, the complementary of unbalanced accuracy. "Loc err" is the number of malpositions above a certain threshold expressed in milliseconds, it is a measure of dispersion of estimates around targets. "Val loss" is the value of the optimization function for the validation set.*

The same approach that characterized the process of implementation of the PPG neural network, was followed also for the design of the PCG one for the detection of S1 sounds.

4.1.2.1 Main architecture

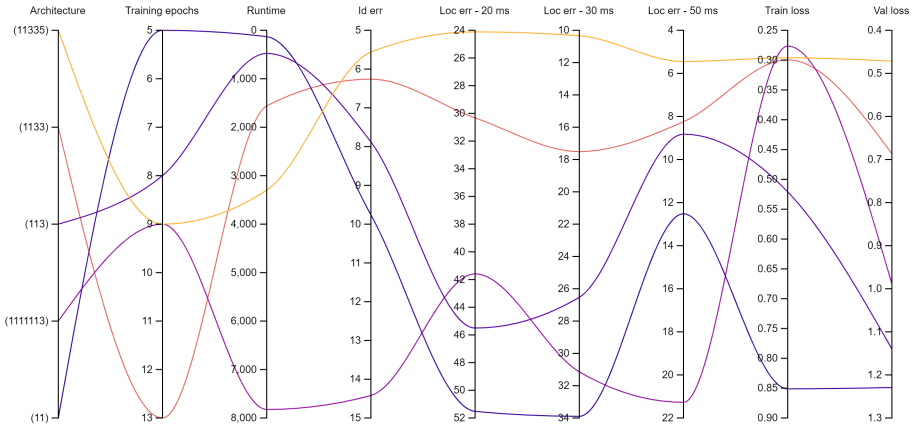

Figura 4.7: *Results of some architecture experiments done for object detection in PCG signals. Each architecture is represented by a vector of numbers of variable length. While the length represents the number of inner blocks constituting the network, the values represent the number of convolutional layer in each block. Each block is implicitly followed by a maxpooling layer.*

First of all, several architectures were tested. In figure 4.7 are reported the results of some of the experimented structure, except for the 113355 one, that inevitably led

to an excessive overfitting on the training set. Again, while simplest networks were not able to learn enough, 11335 and 1133 showed promising outcomes. Surprisingly, even for this task the best network was the 11335, that was selected to be opportunely tuned.

4.1.2.2 Tuning

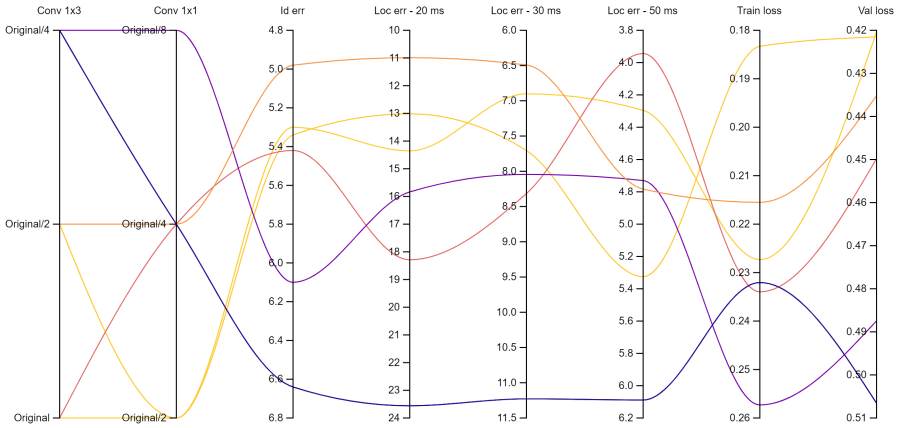

Figura 4.8: *Results of some architecture experiments done for object detection in PCG signals. "Conv 1×1" is the number of 1-sized convolutional layers with respect to the original DarkNet19. "Conv 1×3" is the number of 3-sized convolutional layers with respect to the original DarkNet19.*

Analyzing figure 4.8, it is evident that even in this occasion reducing the number of filters was the best choice. In particular high identification accuracy and low localization error were obtained by halving the number of convolutional filters and further halving the number of 1-sized ones. It can be seen that an excessive reduction in their number, although improved the simplicity of the net, led to worse performance.

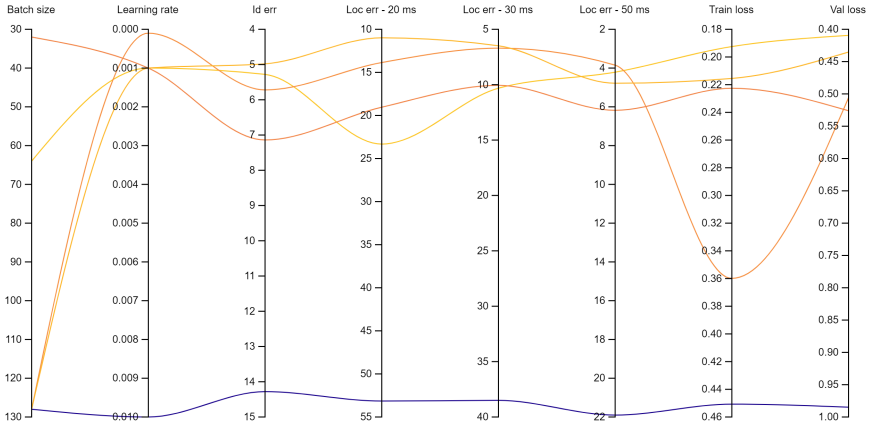

Figura 4.9: *Results of some architecture experiments done for object detection in PCG signals. "Batch size" is the dimension of the batches of samples used fore training. "Learning rate" is the weight given to the prediction error during back-propagation.*

As for the learning parameters, low learning rate and big batches were correlated to better outcomes for all the metrics (Figure 4.9).

4.1.2.3 Input Channels

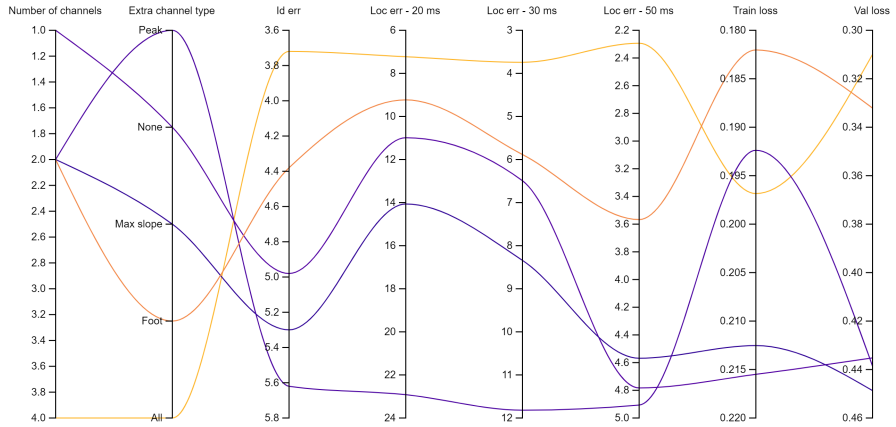

Figura 4.10: *Results of some architecture experiments done for object detection in PCG signals. "Number of channels" is the total number of concurrent input channels. "Extra channel type" is the type of input channel combined with the PCG signal sample, namely the labels that constitutes the outputs of the PPG neural network.*

In figure 4.10 are shown the results obtained by varying the number of input channels. Besides the PCG signal samples, the network was fed also with binary signals indicating the position of the labels in PPG samples. Interestingly, the performance significantly improved especially when the network received as input 4-channel samples, composed by PCG segments and the corresponding binary signals indicating the position of PPG foot, maximum slope and peak in the same segment. These labels were used by the net to improve the performance, by using them as a reference to locate the S1 sounds. This success enabled the possibility to adopt a chain approach, using the output of the PPG net as input for the PCG one. This solution allowed an efficient exploitation of the information coming from the first link in the chain, by converting it in readily-available knowledge.

4.1.3 Final network

In Table 4.1 is reported the final network architecture, actually the same for both signal types. The main difference between the two is the number of input and output channels. While $C_{\text{Input, PPG}}$ is equal to 1, $C_{\text{Output, PPG}}$ is equal to 3, $C_{\text{Input, PCG}}$ is equal to 4, $C_{\text{Output, PCG}}$ is equal to 1.

Type	Filters	Size	Stride	Output
Convolutional	16	1×3	1	600
Maxpool		1×2	2	300
Convolutional	32	1×3	1	300
Maxpool		1×2	2	150
Convolutional	64	1×3	1	150
Convolutional	16	1×1	1	150
Convolutional	64	1×3	1	150
Maxpool		1×2	2	75
Convolutional	128	1×3	1	75
Convolutional	32	1×1	1	75
Convolutional	128	1×3	1	75
Maxpool		1×2	2	37
Convolutional	256	1×3	1	37
Convolutional	64	1×1	1	37
Convolutional	256	1×3	1	37
Convolutional	64	1×1	1	37
Convolutional	256	1×3	1	37
Maxpool		1×2	2	18
Flatten				4608
Dense				$5 \times 2 \times C_{output}$

Tabella 4.1: *Best performing architecture for both the PCG case and the PPG case.*

4.1.4 Model evaluation

In this section are reported the outcomes of the evaluation of the selected networks on each type of label and on different test sets. Table 4.2 shows the custom metrics obtained on the test set distinguishing by type of network and type of labeling. The identification error is basically the complementary of the unbalanced accuracy, while the localization errors indicate haw many labels in percentage were located above a certain threshold expressed in milliseconds. Keeping in mind that in this case the nets' predictions were compared against an automatic labeling algorithm, it is worth noting the profoundly different performance quality according to the label type. In fact, excellent results were obtained in S1 detection and foot detection, while especially in the peak case the outcomes were worse.

Evaluated network	Evaluated label	Ident. error	Loc. error 30 ms	Loc. error 50 ms
PPG	Foot	3.28%	2.91%	1.04%
PPG	Max Slope	5.90%	12.29%	4.76%
PPG	Peak	15.34%	45.64%	26.46%
PPG	All	8.17%	18.40%	9.59%
PCG	S1	2.88%	5.47%	2.16%
PPG+PCG	S1	3.06%	5.28%	2.02%

Tabella 4.2: *Resulting custom metrics after evaluation on test set of the final neural network. For the multi-output PPG net are reported also the results on each single predicted label, while PCG is evaluated with both true PPG labels as input and predicted ones.*

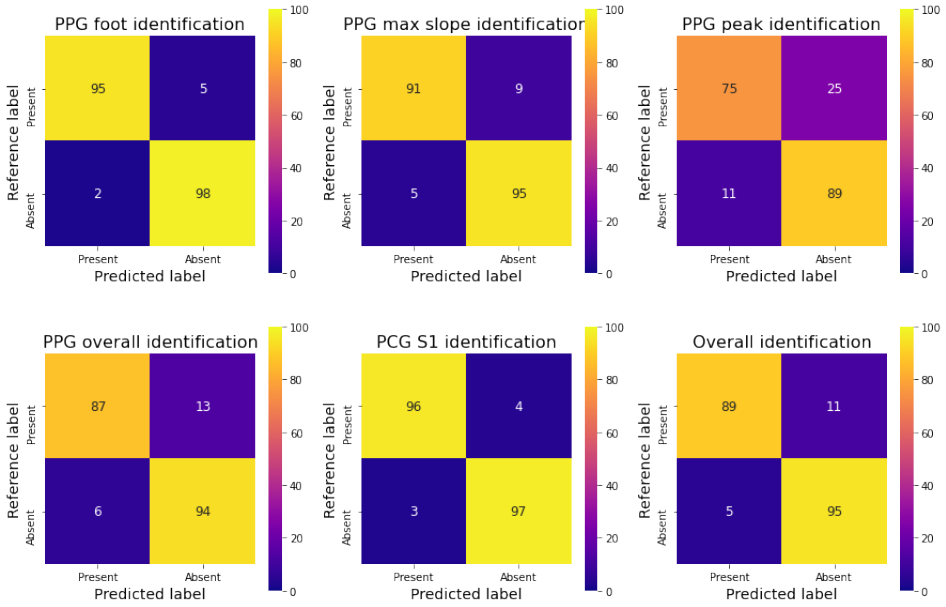

Figura 4.11: *Confusion matrices for the identification task on the test set.*

Going deeper into the results, Figure 4.11 shows the normalized confusion matrices for the identification task and Table 4.3 and 4.4 the standard metrics for binary classification. It can be emphasized that the CNN has similar values for precision and recall. This means the the number of false negatives is close to the number of false positives, so the network outputs a correct absolute number of identifications, but sometimes the reference point is not detected in the correct window. A manual inspection confirmed this hypothesis for both the maximum slope and the peak identifications. Indeed, considering only the identification ability of the network by using a reasonably larger tolerance for localization error, it turned out that the manually estimated balanced accuracy was around 98% per each label type.

Label type	Accuracy	Bal. Acc.	Specificity
PPG_{Foot}	0.97	0.97	0.98
$PPG_{Max\ Slope}$	0.94	0.94	0.95
PPG_{Peak}	0.85	0.85	0.89
PPG_{All}	0.92	0.92	0.94
PCG	0.97	0.97	0.97

Tabella 4.3: *Standard metrics for binary classification evaluating the identification results on the test set.*

Label type	Precision	Recall	F1 score
PPG_{Foot}	0.95	0.95	0.95
$PPG_{Max\ Slope}$	0.90	0.91	0.91
PPG_{Peak}	0.75	0.75	0.75
PPG_{All}	0.86	0.87	0.87
PCG	0.94	0.96	0.95

Tabella 4.4: *Standard metrics for binary classification evaluating the identification results on the test set.*

In Table 4.12 is shown the absolute and the root mean squared error for the localization task, expressed in milliseconds with respect to the target, the output of automatic detection algorithm.

Label type	MAE [ms]	RMSE [ms]
PPG_{Foot}	2.83±0.33	12.30±9.02
$PPG_{Max\ Slope}$	4.49±0.41	15.59±9.22
PPG_{Peak}	8.96±0.69	26.40±9.58
PPG_{All}	5.43±0.29	19.08±7.06
PCG	3.66±0.46	17.08±10.80

Tabella 4.5: *Metrics evaluating the localization results on the test set. Error is reported with its standard deviation.*

Label	Wilcoxon test	Kruskal-Wallis test	Kolmogorov-Smirnov test
Foot	3.84	0.88	0.10
Max slope	5.81	0.75	0.51
Peak	3.42	0.47	0.02
S1	1.24	0.91	0.99

Tabella 4.6: *P-values of statistical hypothesis tests for localization task on test set.*

In figures 4.12-4.13-4.14-4.15 are reported the scatter plots, the Bland-Altman plots and the box plots for each label type, that compare the estimate results with respect to the target.

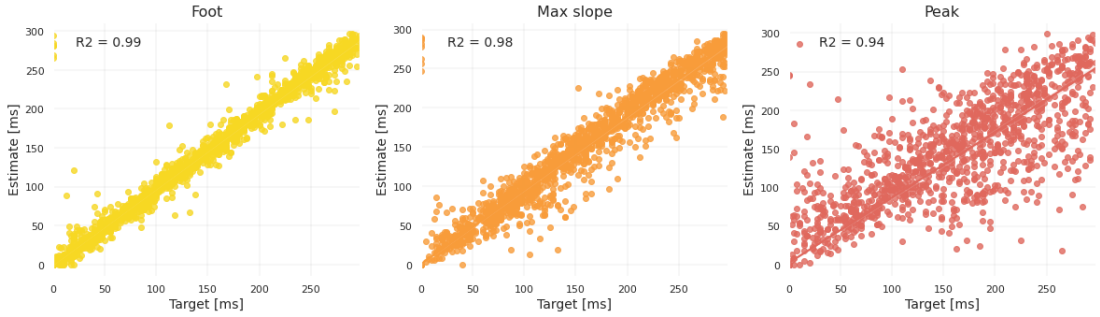

Figura 4.12: *Regression plot for localization task on test set.*

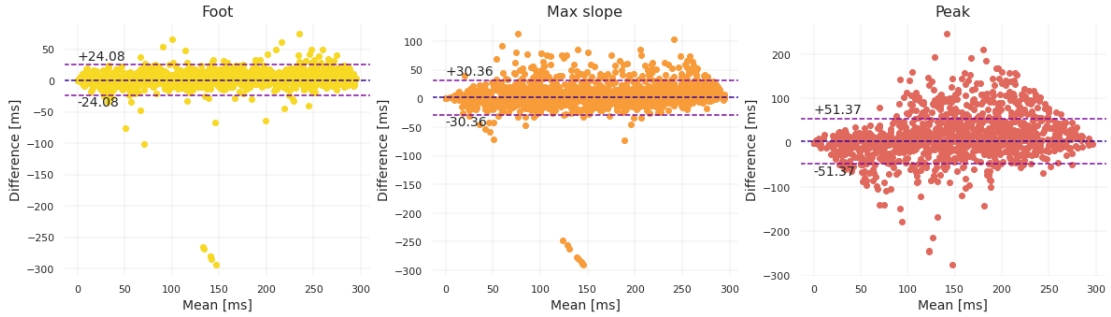

Figura 4.13: *Bland-Altman plot for localization task on test set as a function of estimate position and target position.*

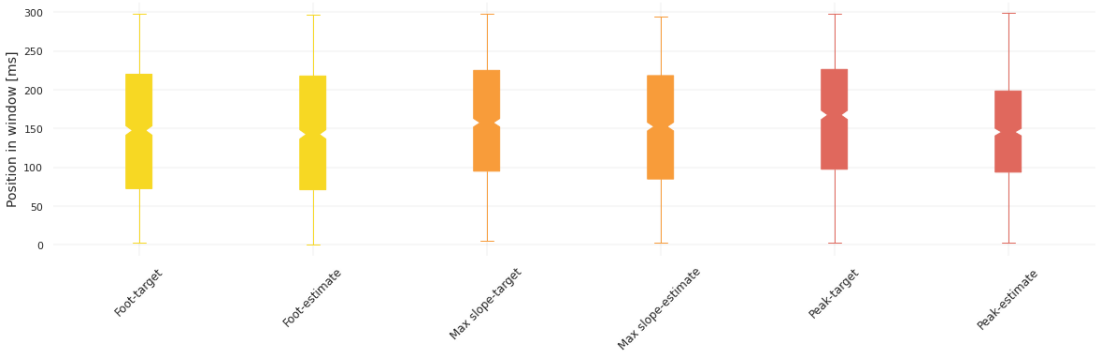

Figura 4.14: *Box plot for localization task on test set.*

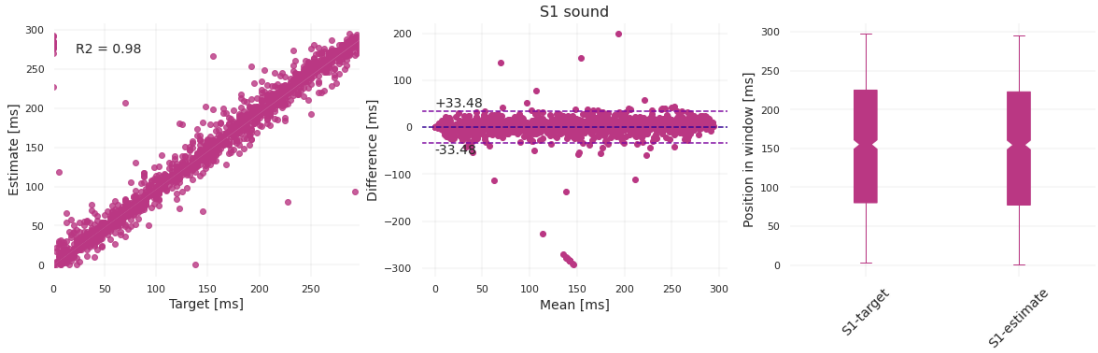

Figura 4.15: *Regression plot, Bland-Altman plot and box plot for localization task on test set.*

4.1.5 Noise sensitivity test

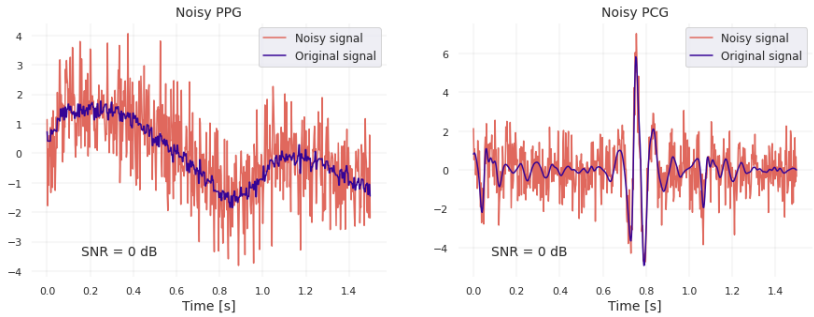

Figura 4.16: *Example of signals affected by added noise.*

In order to investigate the robustness of the networks, a noise sensitivity test was performed. Results are shown in Figures 4.17 and 4.18, where are drawn network performances for the foot case and the S1 case, respectively.

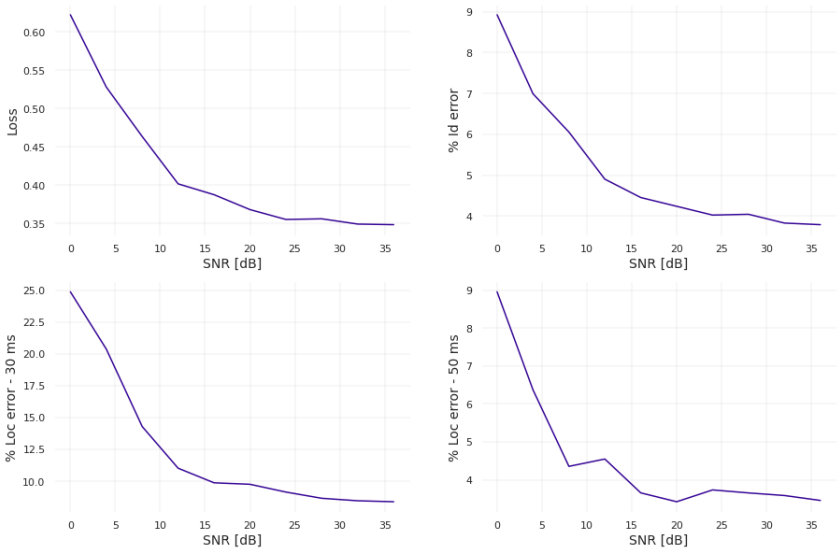

Figura 4.17: *Results of noise sensitivity test for PPG foot detection on test set.*

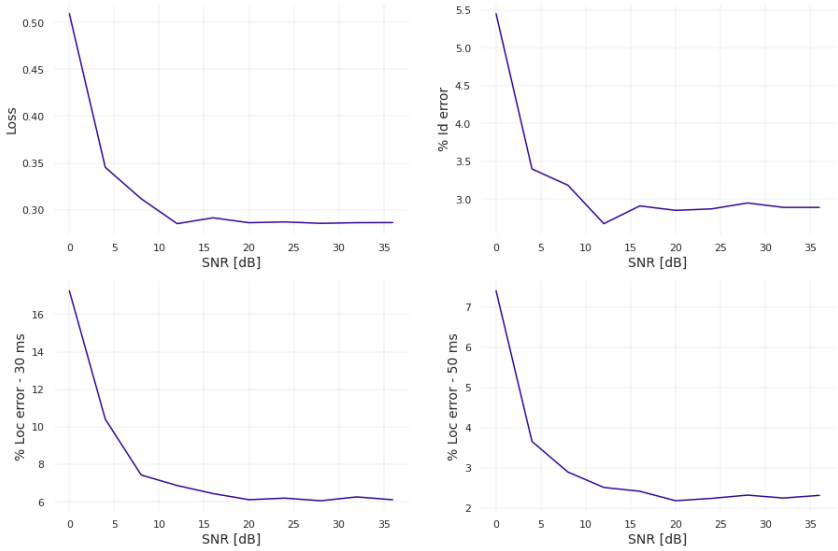

Figura 4.18: *Results of noise sensitivity test for PCG S1 peak detection on test set.*

4.2 Part II: feature extraction from predictions

In part II the detected reference points were used to estimate vascular transit time and heart rate on the test set. Histograms and box plots for VTT are shown in Figure 4.19 and Figure 4.20, respectively; histograms and box plots for HR are shown in Figure 4.21 and Figure 4.22, respectively;

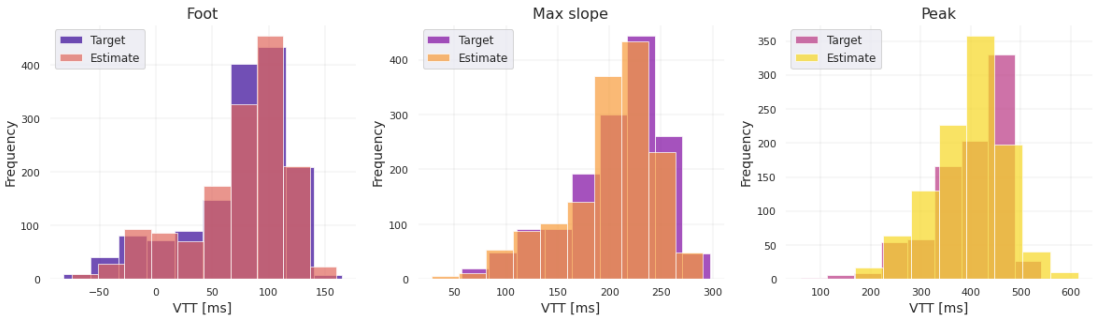

Figura 4.19: *Histograms comparing VTTs computed by using target labels against VTTs computed by using estimate labels, according to the type of label used, on the test set.*

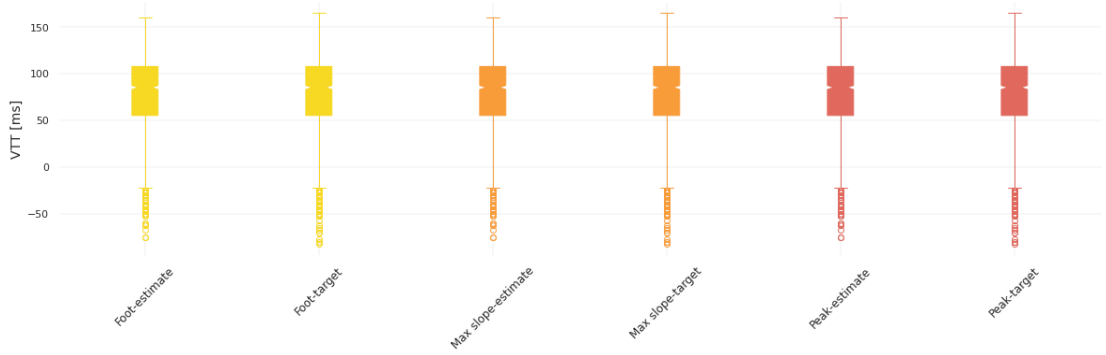

Figura 4.20: *Box-plots comparing VTTs computed by using target labels against VTTs computed by using estimate labels, according to the type of label used, on the test set.*

Label	Kruskal-Wallis test	Kolmogorov-Smirnov test
Foot	0.71	0.78
Max slope	2.27	2.35
Peak	0.06	0.00

Tabella 4.7: *P-values of statistical hypothesis tests for comparing VTTs computed by using target labels against VTTs computed by using estimate labels, according to the type of label used, on the test set.*

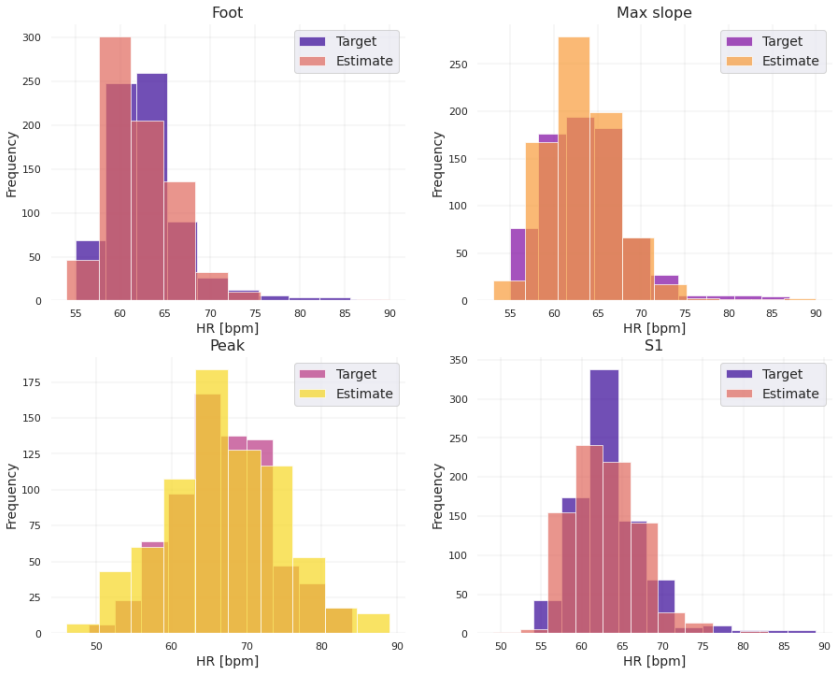

Figura 4.21: *Histograms comparing HRs computed by using target labels against HRs computed by using estimate labels, according to the type of label used, on the test set.*

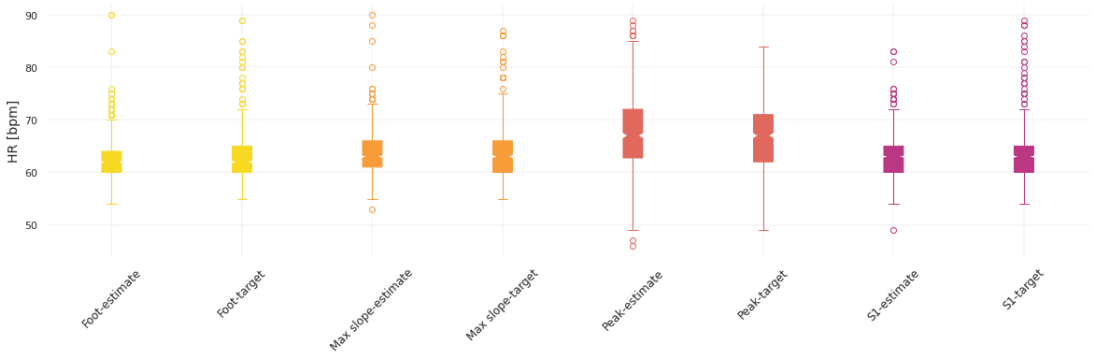

Figura 4.22: *Box plots comparing HRs computed by using target labels against HRs computed by using estimate labels, according to the type of label used, on the test set.*

Label	Kruskal-Wallis test	Kolmogorov-Smirnov test
Foot	0.25	0.82
Max slope	0.27	0.74
Peak	0.34	0.05
S1	0.69	0.95

Tabella 4.8: *P-values of statistical hypothesis tests for comparing HRs computed by using target labels against HRs computed by using estimate labels, according to the type of label used, on the test set.*

4.3 Part III: BP estimation through VTT

4.3.1 Model evaluation

Part III is dedicated to the analysis of the regression models for the estimation of the blood pressure by leveraging the extracted parameters, namely the VTT and the HR. In Table 4.9 are reported the custom metrics on the subject-specific dataset. Since the signals where recorded in less controlled conditions with respect to the datasets used for training and testing, they were also more noisy. Furthermore, recordings happened after very intense physical activity, hence they presents very high HR, although the network was trained with far lower HR. For this reasons the performance of the network was slightly worse than when evaluated on the test set. In Figures 4.23-4.20-4.25-4.26 is reported the statistical analysis for the extracted features as a function of the different labels available. Each feature point is the result of an average over a time window of 30 seconds. It can be seen that the features extracted in this way by using the trained neural networks are very robust and significantly comparable (Tables 4.10-4.11) to the ones extracted with the standard algorithm.

Evaluated network	Evaluated label	Ident. error	Loc. error 30 ms	Loc. error 50 ms
PPG	Foot	7.41%	7.02%	2.79%
PPG	Max Slope	7.66%	13.19%	4.54%
PPG	Peak	14.66%	34.35%	16.41%
PPG	All	8.56%	21.19%	10.82%
PCG	S1	5.50%	6.55%	1.92%
PPG+PCG	S1	6.50%	5.66%	2.05%

Tabella 4.9: *Resulting custom metrics after evaluation on patient-specific dataset of the final neural network. For the multi-output PPG net are reported also the results on each single predicted label, while PCG is evaluated with both true PPG labels as input and predicted ones.*

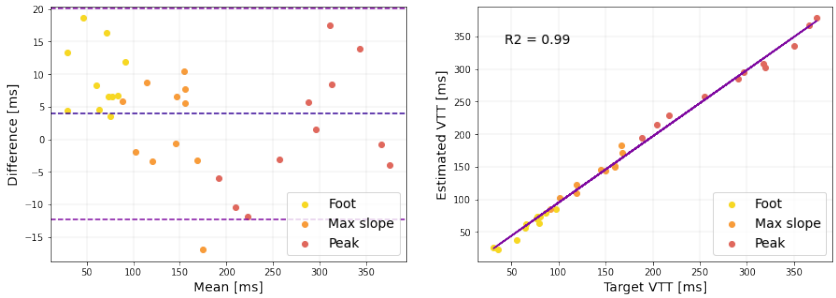

Figura 4.23: *Regression plot and Bland-Altman plot comparing VTTs computed by using target labels against VTTs computed by using estimate labels, according to the type of label used, on the patient-specific dataset.*

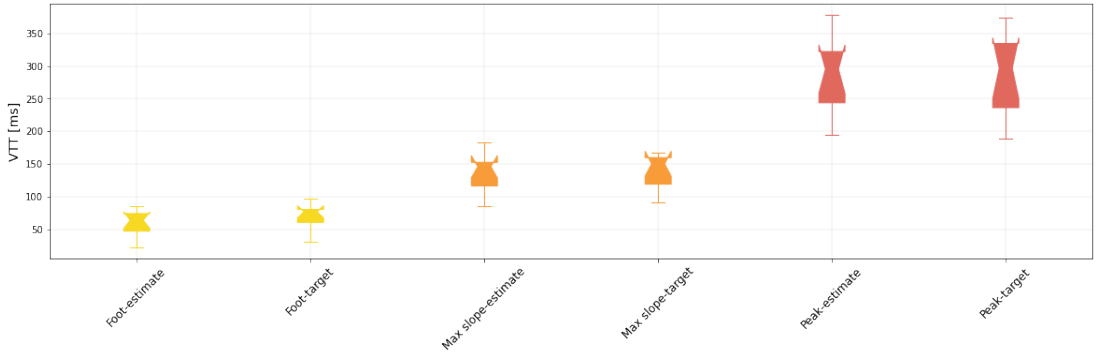

Figura 4.24: *Box plots comparing VTTs computed by using target labels against VTTs computed by using estimate labels, according to the type of label used, on the patient-specific dataset.*

Label	Wilcoxon test	Kruskal-Wallis test	MAE [ms]	RMSE [ms]
Foot	0.04	0.23	9.17±3.42	10.33±8.68
Max slope	0.33	0.81	6.46±3.06	7.79±7.44
Peak	0.95	1.02	7.56±3.61	9.13±8.09

Tabella 4.10: *P-values of statistical hypothesis tests and metrics for comparing VTTs computed by using target labels against VTTs computed by using estimate labels, according to the type of label used, on the patient-specific dataset. Error is reported with its standard deviation.*

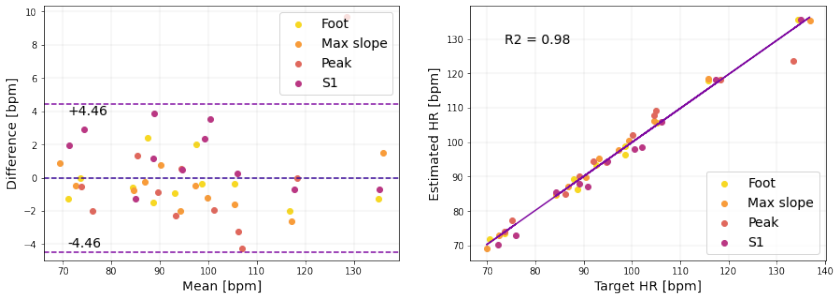

Figura 4.25: *Regression plot and Bland-Altman plot comparing HRs computed by using target labels against HRs computed by using estimate labels, according to the type of label used, on the patient-specific dataset.*

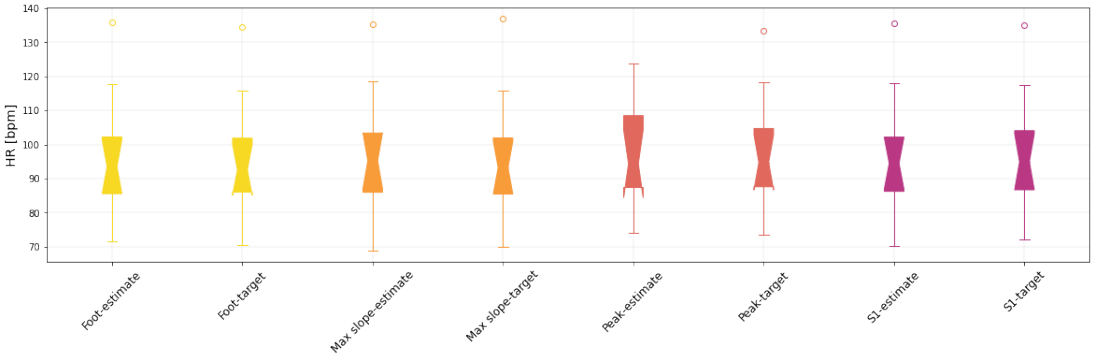

Figura 4.26: *Box plots comparing HRs computed by using target labels against HRs computed by using estimate labels, according to the type of label used, on the patient-specific dataset.*

Label	Wilcoxon test	Kruskal-Wallis test	MAE [bpm]	RMSE [bpm]
Foot	0.36	0.82	1.15±0.51	1.37±1.13
Max slope	0.21	0.93	1.14±0.49	1.33±1.18
Peak	0.22	0.84	2.43±1.81	3.54±4.29
S1	0.14	0.86	1.73±0.85	2.11±1.89

Tabella 4.11: *P-values of statistical hypothesis tests and metrics for comparing HRs computed by using target labels against HRs computed by using estimate labels, according to the type of label used, on the patient-specific dataset. Error is reported with its standard deviation.*

Model	MAE [mmHg]	MSE [mmHg^2]	R^2
1/mean(VTT)	1.81	5.30	0.82
$1/VTT_{\text{max slope}}$	1.81	5.37	0.81
$\log(VTT_{\text{max slope}})$	1.91	5.60	0.81
PCA(1/VTT)	1.97	5.74	0.81
1/VTT	2.09	6.28	0.79
$VTT_{\text{max slope}}$	2.08	6.34	0.78
HR+$1/VTT_{\text{max slope}}$	2.11	6.49	0.78
HR+$\log(VTT_{\text{max slope}})$	2.07	6.56	0.78
log(mean(VTT))	1.99	6.42	0.78
$\log(VTT_{\text{foot}})$	2.32	6.92	0.76
$1/VTT_{\text{foot}}$	2.47	7.57	0.75
$1/VTT_{\text{peak}}$	2.40	7.12	0.75
PCA(HR+1/VTT)	2.34	8.19	0.75

Tabella 4.12: *Simple and multiple linear regression models together with their prediction performance metrics. When a subscript is not explained for VTT, it means that every computed VTT value was included in the model.*

4.3.2 The regression models

The extracted features were then used as inputs for several simple and multiple regression models. The predictions of each model were evaluated as mean absolute error, mean squared error and coefficient of determination with respect to the reference measurements collected through a medical, cuff-based, pressure device. To make the most of the limited amount of available BP samples, a leave-one-out cross-validation approach was adopted. Despite several models were trained for both systolic and diastolic blood pressure estimation, no correlation at all was found for DBP estimations. On the other hand, many significant models for SBP estimation were found. In Table 4.12 are reported the most statistically significant results having a determination coefficient greater than 0.75. When a subscript is not explained for VTT, it means that every computed VTT value was included in the model.

The four models fitting better on the dataset where further analyzed, as shown in Figures 4.29-4.27-4.28-4.30 and are proposed in full in the following.

- Model A: $SBP = a \times 1/VTT_{maxslope} + b$
- Model B: $SBP = a \times \log(VTT_{maxslope}) + b$
- Model C: $SBP = a \times \left(\frac{1}{3}\sum_{i\in(foot,maxslope,peak)} VTT_i\right)^{-1} + b$
- Model D: $SBP = \sum_{i=1}^{N_{PC}} a_i \times PC_i(VTT_{foot}^{-1}, VTT_{maxslope}^{-1}, VTT_{peak}^{-1}) + b$

	Model			
	A	**B**	**C**	**D**
Wilcoxon test	1.00	0.78	1.00	0.92

Tabella 4.13: *P-values of statistical hypothesis test for comparing estimate SBP with respect to SBP, according to the model used, on the patient-specific dataset.*

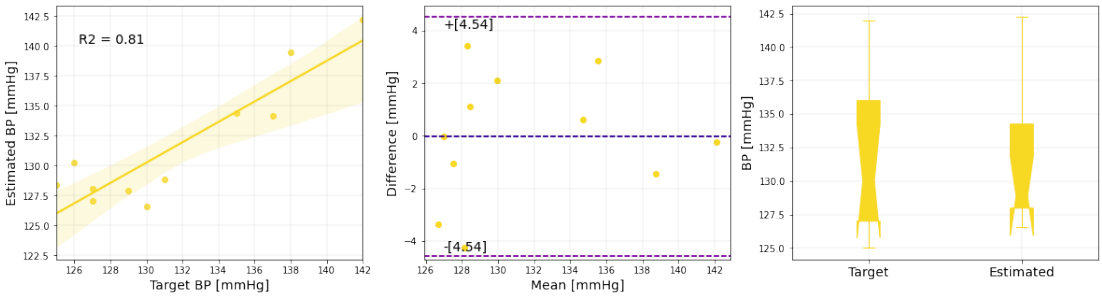

Figura 4.27: *Model A: regression plot, Bland-Altman plot, box plot.*

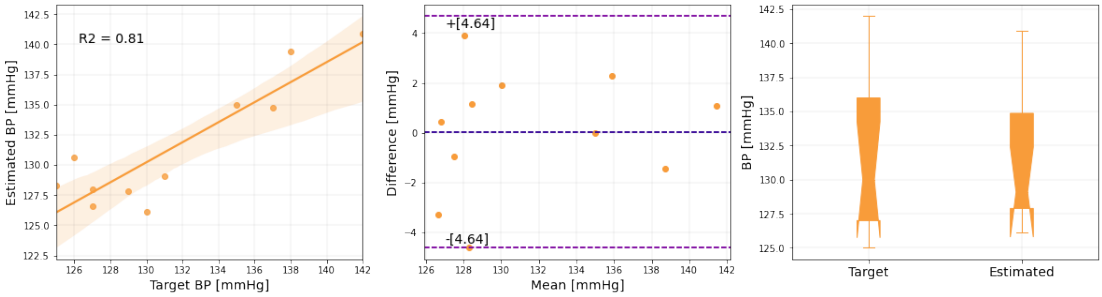

Figura 4.28: *Model B: regression plot, Bland-Altman plot, box plot.*

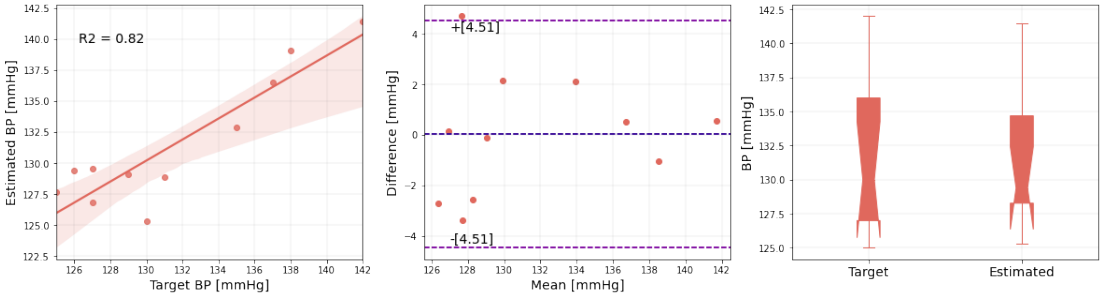

Figura 4.29: *Model C: regression plot, Bland-Altman plot, box plot.*

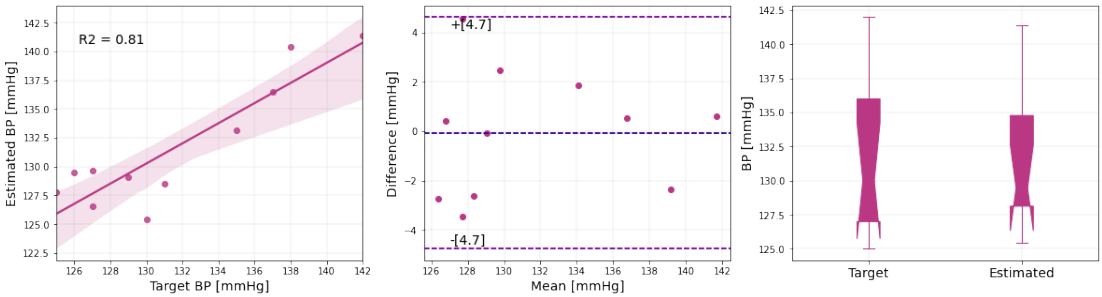

Figura 4.30: *Model D: regression plot, Bland-Altman plot, box plot.*

Capitolo 5

Discussion

5.1 Main findings

5.1.1 Part I: the Convolutional Neural Network

Clinical expert humans glance at a signal and instantly know where useful points of interests are. The human visual system, fast and accurate, together with human knowledge and medical understanding, allows to perform complex tasks like detection of pathological patterns in signals. Detection abilities imply two subtasks: identification of an event and localization of the same along a signal waveform. Usually deep-learning-based object detection aims at recognize an object and its position in 2D images. These tasks are usually accomplished separately, by first classifying the object and then by understanding its position in the frame through a laborious post-processing step. These complex pipelines are slow and hard to optimize because each individual component must be trained separately. In this thesis project the separate components of object detection have been unify into a single, powerful, fast convolutional neural network optimized to process specifically one-dimensional signals. The object detection problem has been reframed as a single hybrid regression-classification problem, straight from signal samples to event coordinates and identification probabilities.

The overall network architecture is composed by 13 convolutional layers followed by 1 fully connected layer. It can be conceptually divided in two parts with different purposes. The first part is composed by the sequence of convolutional layers interspersed by maxpooling layers. This substructure behaves as future extractor: its aim is to identify characteristic patterns in data, in order to recognize the presence of events of interest, such as the reference points used in the project, namely the PPG foot, maximum gradient and peak as well as the PCG S1 peak. The number of convolutional layer and filters progressively increase in each block going deeper in the net. This configuration has been adopted to induce the network to build increasingly complex features at deep layers by combining the more simple ones extracted at superficial layers. Along the net, feature space compression has been done to discourage overfitting and sustain generalization, by using maxpooling layers and, in a more peculiar way, by exploiting 1-sized convolutional layers. In fact, alternating this convolutional layers to the 3-sized ones reduce the features space from preceding layers, by combining features across all the feature maps. Extracted features are then processed by the final dense layer, con-

stituting the second and last part of the network, that provides the output. Essentially it is responsible for the localization of the identified objects. The output shape is designed so that the input signal segment is divided in five equidimensional windows. If an object falls into a window, that window is responsible for detecting that object. Each window is associate to two output neurons for each label type: one expressing the normalized position of the label inside the segment, the other a confidence scores for this identification. This confidence score reflects how confident the model is that the segment contains an object. If no object exists in that segment, the confidence scores should be zero, otherwise the confidence score should be one. The final fully-connected layer can have a variable dimension according to the number of objects to be identified, each one constituting a different output channel C_{output}. Its purpose is to guarantee extreme flexibility about the number of simultaneous detected labels. It has been proved that the network performance is not affected by the number of output channels, since its performance on each label was comparable when tested on multiple-object with respect to single-object case. This fact means that the features extracted by the first part of the net are general enough to describe entirely the morphological characteristics of signal segment provided as input. The possibility to use the same network, at the same time, with the same computational cost, to accomplish different tasks makes it valid and efficient for different purposes in the most disparate fields and contexts. Staying in the biomedical field, the designed and implemented algorithm can be exploited for example to detect events in electrocardiographic signal, such as the QRS complex and its elementary points, in ballistocardiographic signal, whose points could also be use as a proxy to compute the pulse arrival time, in EEG, to identify spikes, etc. Furthermore, it has been shown the possibility to feed the network with a multi-channel input segment, where each channel C_{input} represents a synchronous signal segment. It has been demonstrated also how effective this approach could be: in fact, in chain configuration, when the network for S1 recognition was fed with detected labels of PPG, the accuracy of prediction and localization improved significantly. This approach has also been shown very robust to noise in input data, even when PPG labels were significantly unreliable. This expedient allows to effectively exploit the readily-available information coming from the PPG network to improve the ability to discriminate between S1 and S2 sounds, that could eventually be very similar, especially in noisy signals. The possibility to use the same network architecture with a different number of input channels opens up a variety of possibilities and use cases. For instance, the network can be fed with a set of wavelet components previously crafted, as well as frequency content and time content together. Point detection is a simple task, but could be the key to enable several more complex tasks, like, for example, signal segmentation. Once one knows useful fiducial points in a signal, it is easier to perform its automatic segmentation.

Going into the results of the evaluation on the test set, it can be seen that excellent results were obtained especially for foot and S1 detection. In fact, in the identification task, both the PPG network and the PCG one achieved 97% of balanced accuracy on the cited labels, while on maximum slope detection it was 94%. As for the localization error, Bland-Altman shows an almost-zero mean difference between estimate and target, as well as limited dispersion of data around the mean difference (1.96*sd $\approx$ 30 ms). Also, the regression plot shows a coefficient of determination very close to 1. A

definitely worse performance was achieved on PPG peak detection. Balanced accuracy was 85% and, despite the mean of the Bland-Altman's difference was again almost zero, the dispersion was doubled. A deeper analysis of the statistics shows a fundamental insight: the CNN has similar values for precision and recall. This implies that the number of false negatives is equal to the number of false positives, thus the network outputs a correct absolute number of identifications even though sometimes in a wrong window. A manual inspection over almost 200 beats validated this hypothesis for every label type. Computing the identification accuracy without considering correct only the specific window, but also the contiguous ones, allows to adjust the estimated balanced accuracy for identification around 98% per each label type. These data highlight the fact that the problem is not the identification ability of the network, but rather its localization skill. The reason of this skill degradation must be sought in two influencing factors: the intrinsic functioning of convolutional neural networks and the labeling quality. As for the functioning, it can be said that the extraction of high level features inevitably leads to the loss of low level information, thus the details useful to precisely identify specific points, increasing the complexity of the localization. A possible approach to solve this issue is proposed in the Section 5.4. As for the labeling quality, on a noisy PPG morphology feet are easier identifiable with respect to maximum slopes and peaks. In fact, while in high quality signals the systolic peak is higher than and distinct from the diastolic one, on real-world data it can be lower than the diastolic one, as well as merged with it. Therefore, the automatic labeling algorithm developed in this project, although quite robust, was not always able to correct this localization error. These errors constituted noise in the training set that led the network to find a local minima in its loss function in correspondence of an intermediate point between the systolic peak and diastolic one. Anyway, the robustness of the algorithm is evident when the network is fed with correct samples.

Another characteristic of the network that deserves to be explored is how the segments are processed to perform a detection. It is worth noting that, even though a positive prediction occurs when an object is in a specific window, the network processes the whole segment to perform the task. This means that it uses features from the entire segment to predict, accounting also for the feature-encoded presence of contiguous labels. It is able to detect more objects simultaneously by reasoning globally about the full segment and all the present objects. Somehow, it is probably able to keep in consideration also the contextual information such as the heart rate, since the cardiac period is the distance between labels of the same type. Adaptability to HR modifications have been demonstrated by evaluating the network on the dataset used in Part III (Table 4.9). Despite the network was not trained on signals with such high HR, it was able to correctly detect labels in the correct window in about 93% of samples. Anyway the true identification performance is considerably better, since the manual inspection reveals that identification accuracy without considering the specific window was about 97%, a result consistent with the one obtained through the manual inspection on test set.

Another point that requires closer examination is the pre-processing of the input signals the network can work with. In this project have been shown the possibility to work even with almost raw signals. In fact, while PCG was remove of its low frequency to recover the characteristic waveform, the PPG was not filtered at all. The only requi-

red step was a normalization of each segment by removing the mean and bringing the variance to one, an expedient that facilitates the network reasoning. Indeed, it ensures the presence of both positive and negative values used as inputs for the next layer, which makes learning more flexible, and that the network learning regards all input features to a similar extent. This light processing is of great value, because reduces computational cost and time resources needed to perform the detection. Furthermore, the presence of noise in the training data increase the robustness of the network acting as a regularization factor by preventing the overfitting and improving the generalization ability. To further exaggerate these positive effects, noise was also artificially augmented by introducing different kinds of disturbing signals, namely uniform noise, Gaussian noise and periodic noise. In order to test the resistance to noise of the network, a noise sensitivity test was performed. As shown in Figure 4.17 and Figure 4.18, the neural network demonstrate a surprising robustness to Gaussian noise, especially in the identification task.

5.1.2 Part II: feature extraction from predictions

In order to compare the distributions of the features extracted by reference prediction and the network ones, two statistical test were performed: two-sample Kolmogorov-Smirnov test and Kruskal-Wallis test. Kolmogorov-Smirnov test (KS test) is a non-parametric test of the equality of continuous, one-dimensional probability distributions that can be used to compare two samples by statistically quantifying the distance between the empirical distribution functions of two samples. The null distribution of this statistic is calculated under the null hypothesis that the samples are drawn from the same distribution. The two-sample KS test is one of the most useful and general non-parametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples. The Kruskal-Wallis test (KW test) by ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of equal or different sample sizes. The parametric equivalent of the KW test is the one-way analysis of variance (ANOVA). A significant KW test indicates that at least one sample stochastically dominates one other sample. Since it is a non-parametric method, it does not assume a normal distribution of the residuals, unlike the analogous one-way analysis of variance.

For both VTT and HR, for all labels, KW test showed that the differences between the medians were not statistically significant. Furthermore, two-sample KS test showed that also the differences between the empirical distribution functions were not statistically significant, except in one case. Predictably, $\mathrm{VTT}_{\mathrm{peak}}$ was affected by the weighty error of localization, whose presence was justified in the previous section. Two insights can be gained from these outcomes. First of all, results confirm the general possibility to use the network extracted labels to estimate both vascular transit time and heart rate, boosting the reliability of Part III, whose results are entirely based on these features. Secondly, results confirm the general validity of the feature computation method implemented.

5.1.3 Part III: BP estimation through VTT

Since in Part III the extracted features were matched with the target ones, besides the KS test and the KW test, also a Wilcoxon test was performed. The Wilcoxon signed-rank test is a non-parametric statistical hypothesis test used to compare two matched samples to assess whether their population mean ranks differ. It can be used as an alternative to the paired Student's t-test when the distribution of the differences between the two samples cannot be assumed to be normally distributed. It determines whether two dependent samples were selected from populations having the same distribution. All the tests, for both VTT and HR, for all the labels, except Wilcoxon test on PPG foot, do not refused the null hypothesis claiming that the populations were the same, thus the differences between samples were not statistically significant. Analyzing deeper the foot case, a curious behavior emerges: an evident overestimation of $\mathrm{VTT}_{\mathrm{foot}}$ occurs as shown in the Bland-Altman plot, but $\mathrm{HR}_{\mathrm{foot}}$ seems to be consistent with the estimations done employing the other labels. It can be hypothesized a systematic and constant overestimation of the foot position, but more data would be needed to have reliable results.

As for the systolic blood pressure estimation models, as expected the maximum slope label and the foot label were the most useful for BP estimation. Since they were less affected by localization error, more reliable features were extracted with respect to when peak reference point was used. The four most promising models among the tested ones showed an evident linear correlation between estimate and target, with a determination coefficient greater than 0.80, while mean absolute error and mean squared error were less than 2 mmHg and 6 mmHg^2, respectively. For every model, Bland-Altman analysis showed a dispersion around the mean difference lower than 5 mmHg and a null mean difference.

Model A, validated in literature with VTT by Esmaili et al. in [3], that proposes a reciprocal relationship between BP and VTT, here finds confirmation of its value for systolic blood pressure assessment. However, the aforementioned research team found a high correlation coefficient also for the diastolic blood pressure estimation case, while here the determination coefficient after a leave-one-out cross-validation was exactly zero for the same case. In Table 5.1 are reported their results for this model accounting also for the different reference points used. Even though the estimation error obtained in this project is sensibly better than the one obtained by the research team, the dimension of the used support must be taken into account, since their conclusions were inferred from an analysis on 24 subjects. It is worth noting the similar determination coefficients for both foot and maximum slope labels, as well as the difference between the peak-related ones. Very similar results were obtained in this project also with model B, describing a logarithmic relationship between VTT and BP.

Even better prediction outcomes were obtained by exploiting model C that was also the most performing one, having a coefficient of determination equal to 0.82 and consequently slightly lower estimation error. This model, as model A, describes a reciprocal relationship between VTT and BP, but here the time delay is measured as an average of all the computed VTTs, namely $\mathrm{VTT}_{\mathrm{Foot}}$, $\mathrm{VTT}_{\mathrm{Max\ plot}}$ and $\mathrm{VTT}_{\mathrm{Peak}}$. It can be hypothesized that the computation of the mean value partially reduces the noise carried by elementary VTTs, reducing the variance of the extracted feature. This model makes the most of the network capabilities, since it simultaneously exploits every

Label	Study	MAE [mmHg]	R^2
Foot	This work	2.47±2.41	0.75
	[5]	1.14±14.94	0.71
Max slope	This work	1.81±1.36	0.81
	[5]	7.37±11.09	0.84
Peak	This work	2.40±1.15	0.75
	[5]	7.47±11.08	0.84

Tabella 5.1: *Comparison between results obtained by exploiting model A in this project against the ones obtained by Esmaili et al. in [5].*

object detected by the algorithm.

Finally, good results were obtained also leveraging model D. As model C, it exploits every label computed and it is essentially based on the reciprocal on VTT. Anyway, here the combination of the elementary VTTs is done through a principal component analysis transformation. PCA allows to include a reasonably high number of features also when the number of samples is limited without incurring in overfitting, since it is averted by reduction of the feature space by an appropriate selection of the linear combinations of features that explain a fraction of the variance considered satisfying.

5.2 Work limitations

Despite the successful outcomes of this project, several limitations might be highlighted, some of which have already been mentioned before. Here in the following are listed and explained the most crucial of them:

- Labeling accuracy. As shown in previous sections, labeling quality have influenced the network abilities in a not negligible way. While foot, maximum slope and S1 peak were precisely identified and localized even with a semiautomatic labeling algorithm, the systolic peak labeling was affected by the lack of accuracy in detection. The effect was a degeneration of the network detection ability, partially compensated only by the adoption of an average over a time window to estimate features like HR and VTT. Thus, especially for eventual real-time, beat-to-beat applications, a manual labeling seem to be necessary to provide reliable information on which the network can effectively learn.

- Dataset dimension. While the dimension of the dataset used to train the neural network was satisfying, also because of the decision to take a data augmentation step, the dataset used for blood pressure estimation was limited, even though not enough to obtain encouraging results. Of course, this dataset needs to be enhanced by collecting measurements from as many subjects as possible. Furthermore, the higher the heterogeneity of the subjects joining the experiment, the higher will be the robustness and generalization ability of the algorithm implemented for blood pressure estimation.

- DBP model absence. If on the one hand a powerful and encouraging algorithm for systolic blood pressure assessment have been developed, on the other no working

solution have been found to map network-extracted features to diastolic blood pressure. DBP is notoriously less prone to sudden changes [5, 3], so it may be more difficult to obtain a wide range of values for the same subject. Further tests must be executed.

5.3 Unsuccessful attempts

Parallel to what have been presented, another attempt was done to estimate blood pressure by using a deep learning approach. During the recordings of the neural network dataset, synchronized BP measurements were recorded by using a validated (AAMI/ISO 81060-2) sensor produced by Maxim Integrated and embedded on the MAXREFDES220# System Board. According to the company claims, it provides everything one needs to measure finger-based heart rate, blood oxygen saturation level and blood pressure trending. It uses an optic sensor to collect the photopletismographic signal from the fingertip, which is processed in real-time by the integrated algorithm, the Maxim Blood Pressure Trending. The system requires user-specific calibrations through cuff-based measurements, then the BP estimations are provided by analyzing the changes in PPG morphological characteristics relative to the calibration instance. In principle, changes in PPG shape are correlated to BP changes. Some of the determinants of the PPG shape have been described in literature, such as systolic amplitude, pulse width, pulse area, peak-to-peak interval, pulse interval and artery stiffness index.

Figura 5.1: *Logo of Maxim Integrated*

Since the sensor outputs a BP per heart beat, each BP measurement was associated with a set of hand-crafted features extracted leveraging the algorithm explained in Part II. Features were the cardiac period, systolic period, diastolic period, S2-S1 interval, S1-S2 interval, VTT_{Foot}, $\text{VTT}_{\text{Max slope}}$, VTT_{Peak}, systolic time and diastolic time. These features were then used to build a dataset to train a neural network that would map the features themselves to the corresponding values of DBP and SBP. The network tested were simple dense networks made of a couple of layers, regularized by connection dropouts, with linear, final activation function and RMSE loss function. Results were extremely far from expectations, since on independent samples the trained network was not able to follow the blood pressure trends by relying on these features. A deeper analysis of the correlation between features and blood pressure was also done to investigate their relationships. No significant correlation was found between feature variations and concurrent estimated BP changes. Furthermore, intra-patient variation range in both features and BPs was quite limited, surely enough to avoid a generalized learning of the neural network.

5.4 Future development

Network improvements

- Improve the labeling. As previously said, train the network with a more accurate labeling could potentially improve its localization skill. Therefore resources

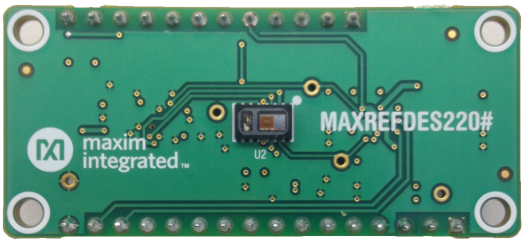

Figura 5.2: *MAXREFDES220# System Board.*

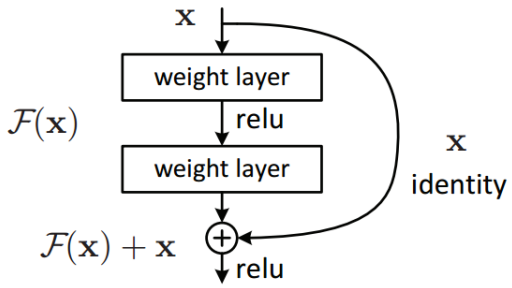

Figura 5.3: *Residual block functioning.*

could be invested in an accurate, manual, labeling process, since the automatic algorithm was not always able to meet expectations, in particular when it had to discern the systolic peak from the diastolic one.

- Introduce residual layers. The extraction of high level features through convolutional and maxpooling layers inevitably leads to the loss of low level information, thus of the details useful to precisely identify specific points, increasing the complexity of an accurate localization. To partially preserve this information, residual blocks can be introduced in the neural architecture. A residual block allows to fast-forward the activation of a layer to a deeper layer. Essentially, residual blocks allows the flow of low-level features from initial layers to final layers, improving the accuracy of deep networks and reducing the vanishing gradient problem that usually affect deep networks.

- Implement partial overlapping. Since each track was divided in smaller segments, it could have happened that some PPG peaks were divided in two distinct segments. In general, this was not a problem, but when the separate tract was the systolic phase, it could happen that two predictions for the same reference point were made in two contiguous segments. In a real-time application perspective, the implementation of a partial overlapping between segments could help solving this kind of error. In this framework, predictions would be weighted more when positioned in a central section of a segment to account for the null probability to have an interrupted tract. Consequently, among the correspondent detections, only the most reliable one would be considered true.

- Improve final block of layers. In the current architecture, flatten layer of features is suddenly fully connected to the output layer. This determines a sudden tran-

sition from more than 4 thousands neurons to just few units. Despite this abrupt transition has been proven effective, more sophisticated configurations prudently managing this step can be tested and eventually adopted. For instance, some of the possible alternatives could be: an intermediate dense layer with some dozens of neurons to smooth the transition, a dropout layer to select only the most influential connections, maxpooling layers to actively contract the feature space.

- Test batch normalization. Batch normalization allows the standardization of the input features of each layer, analogously to what has been done here with input samples of the network. It normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. This solution allows to increase the stability of a neural network, reducing the overfitting on the training set and boosting the learning speed.

- Test a unique CNN for simultaneous labeling of different signals. In this project two networks arranged in a row were used to process two distinct signals, sometimes including a multi-channel input. Anyway, a different approach could be tested: since the two signals are simultaneous, they could be provided as input for the same network as two channels belonging to the same sample. The network would follow the same multi-labeling practice adopted in the current work, but the output would detect labels in different signals at the same time. The adoption of this approach could possibly halving the computational cost and the prediction speed, since the same low-level feature maps would be used to build high-level knowledge.

- Different activation functions for classification and localization. The final activation functions of the implemented networks are linear for both the neurons aimed at identification task and the ones aimed at localization. Although localization is a regression task, thus suitable for this kind of activation, the binary classification could earn benefits from a non-linear activation function such as the sigmoidal one, since it shrinks the output towards 0 or 1 in a more distinct and effective way.

BP estimation improvements

- Improve the database. As already explained, the limited dimension of the dataset was limited, since it was made of a unique subject-specific collection of measurements. Heterogeneous, multi-subject dataset can provide solid, statistical basis to evaluate the goodness of the blood pressure estimation models hypothesized in this project. Furthermore, it can operate a more accurate selection of the models, highlighting the differences in performance between each other.

- Test other BP perturbations. To generate a subject-specific blood pressure estimator, it is imperative to collect a wide range of BP values in order to reliably assess the parameters governing the deterministic models implement. Therefore, perturbations affecting the systolic and diastolic BP are needed to generate numerous and quick changes to be used for model calibration. As shown in Table 2.3, it is full of methods to induce perturbations in blood pressure. Further tests

should be conducted to broaden the understanding of the effects of the different methods, as well as how they affect the two BP types differently.

- Implement a DBP estimation model. Although systolic blood pressure is paramount for monitoring health conditions, also diastolic blood pressure is useful, especially for young subjects. For this reason also a model able to estimate DBP is desirable, in order to complete the overview on the subject's health condition.

General improvements Besides technical improvements specifically affecting the deep learning model and the regression models, there are some improvements working at the general level of the processing pipeline:

- BCG and ECG integration. As seen in the Section 3, several signal types can implicitly carry information about blood pressure level and can be used to extract features to obtain its proxy. Signals such as electrocardiogram and ballistocardiogram could be integrated in the models, leveraging the same architectures, to increase accuracy and robustness of the features extracted and to eventually craft new features.

- Real-time integration. The implemented architecture is flexible and light enough to be applied for real-time monitoring of blood pressure. Potentially, for every processed segment of 1.5 seconds, the time intervals extracted after point detection can be used as inputs for the regression models. In order to keep robust the estimation, a moving average can be employed to account also for preceding estimation and avoid sudden, unreliable changes due to noise on the signals. Furthermore, a mechanism for signal quality assessment could be useful to reduce wrong feature estimations, as well as a filtering step to delete outlier time intervals.

Capitolo 6

Conclusions

In this project has been designed and implemented a fast, flexible, accurate, neural network for object detection in time series. The same network is able to both recognize and localize multiple events simultaneously with almost the same computational cost. This ability to accomplish different tasks at the same time makes it valid and efficient for different purposes in the most disparate fields and contexts, for example to detect QRS complexes in ECG signals. Furthermore, the possibility to use the same network architecture with a different number of input channels opens up a variety of possibilities and use cases. For instance, the network can be used to process concurrent frequency and time contents. Despite point detection is a simple task, it could be the key to reach several more complex targets, like, for example, signal segmentation.

In this thesis, also a use case demonstration of the network potentialities has been provided. The features extracted by using the network, namely vascular transit time and heart rate, have been successfully employed to estimate the systolic blood pressure, leveraging and strengthening the simple physiological, regression models adopted in literature in similar contexts. Future developments include the necessity to test the models on a greater number of subjects, as well as the implementation of a model for diastolic blood pressure. Although further studies are needed to investigate these results, these are very encouraging and consistent with expectations, as well as literature outcomes.

Considering the encouraging obtained results, in future the developed deep learning algorithm can be implemented on Soundi for continuous blood pressure estimation, overcoming the shortcomings of traditional methods.

Bibliografia

[1] Alberto Porro and Enrico Vannoni. *Development of a wearable blood pressure device based on pulse non-invasive measurements*. PhD thesis, Politecnico di Milano, 2017.

[2] Davide Marzorati, Dario Bovio, Caterina Salito, Luca Mainardi, and Pietro Cerveri. Chest Wearable Apparatus for Cuffless Continuous Blood Pressure Measurements Based on PPG and PCG Signals. *IEEE Access*, 8:55424–55437, 2020.

[3] Amirhossein Esmaili, Mohammad Kachuee, and Mahdi Shabany. Nonlinear Cuffless Blood Pressure Estimation of Healthy Subjects Using Pulse Transit Time and Arrival Time. *IEEE Transactions on Instrumentation and Measurement*, 66(12):3299–3308, dec 2017.

[4] Jong Yong Abdiel Foo, Chu Sing Lim, and Ping Wang. Evaluation of blood pressure changes using vascular transit time. *Physiological Measurement*, 27(8):685–694, 2006.

[5] Amirhossein Esmaili Dastjerdi, Mohammad Kachuee, Mahdi Shabany, Amirhossein Esmaili, Mohammad Kachuee, and Mahdi Shabany. Non-invasive blood pressure estimation using phonocardiogram. In *Proceedings - IEEE International Symposium on Circuits and Systems*. Institute of Electrical and Electronics Engineers Inc., sep 2017.

[6] Mohamed Elgendi, Richard Fletcher, Yongbo Liang, Newton Howard, Nigel H. Lovell, Derek Abbott, Kenneth Lim, and Rabab Ward. The use of photoplethysmography for assessing hypertension. *npj Digital Medicine*, 2(1), dec 2019.

[7] Manuja Sharma, Karinne Barbosa, Victor Ho, Devon Griggs, Tadesse Ghirmai, Sandeep Krishnan, Tzung Hsiai, Jung-Chih Chiao, and Hung Cao. Cuff-Less and Continuous Blood Pressure Monitoring: A Methodological Review. *Technologies*, 5(2):21, may 2017.

[8] Chun-Chieh Chieh Hsiao, Joe Horng, Ren-Guey Guey Lee, and Robert Lin. Design and implementation of auscultation blood pressure measurement using vascular transit time and physiological parameters. In *2017 {IEEE} International Conference on Systems, Man, and Cybernetics ({SMC})*, volume 2017-Janua, pages 2996–3001. IEEE, nov 2017.

[9] Joseph Redmon and Ali Farhadi. YOLO9000: Better, Faster, Stronger. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017-Janua:6517–6525, dec 2016.

[10] Bernd Saugel, Ron Dueck, and Julia Y. Wagner. Measurement of blood pressure. *Best Practice and Research: Clinical Anaesthesiology*, 28(4):309–322, dec 2014.

[11] Shanchen Pang, Tong Ding, Sibo Qiao, Fan Meng, Shuo Wang, Pibao Li, and Xun Wang. A novel YOLOv3-arch model for identifying cholelithiasis and classifying gallstones on CT images. *PLoS ONE*, 14(6):e0217647, jun 2019.

[12] World Health Organization. Cardiovascular diseases (CVDs).

[13] Sverre E. Kjeldsen. Hypertension and cardiovascular risk: General aspects, mar 2018.

[14] World Health Organization. Hypertension.

[15] Bryan Williams, Giuseppe Mancia, Wilko Spiering, Enrico Agabiti Rosei, Michel Azizi, Michel Burnier, Denis L. Clement, Antonio Coca, Giovanni de Simone, Anna Dominiczak, Thomas Kahan, Felix Mahfoud, Josep Redon, Luis Ruilope, Alberto Zanchetti, Mary Kerins, Sverre E. Kjeldsen, Reinhold Kreutz, Stephane Laurent, Gregory Y.H. Lip, Richard McManus, Krzysztof Narkiewicz, Frank Ruschitzka, Roland E. Schmieder, Evgeny Shlyakhto, Costas Tsioufis, Victor Aboyans, and Ileana Desormais. 2018 ESC/ESH Guidelines for the management of arterial hypertension. *Journal of Hypertension*, 36(10):1953–2041, oct 2018.

[16] Elena Chung, Guo Chen, Brenton Alexander, and Maxime Cannesson. Non-invasive continuous blood pressure monitoring: A review of current applications, mar 2013.

[17] Karsten Bartels, Stephen A. Esper, and Robert H. Thiele. Blood Pressure Monitoring for the Anesthesiologist. *Anesthesia & Analgesia*, 122(6):1866–1879, jun 2016.

[18] David B. Wax, Hung Mo Lin, and Andrew B. Leibowitz. Invasive and concomitant noninvasive intraoperative blood pressure monitoring: Observed differences in measurements and associated therapeutic interventions. *Anesthesiology*, 115(5):973–978, nov 2011.

[19] John Allen, Crispian P Oates, Timothy A Lees, Al , Alan Murray -, and Kevin Howell -. Physiological Measurement Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas*, 28:1, 2007.

[20] Mohamed Elgendi. On the Analysis of Fingertip Photoplethysmogram Signals. *Current Cardiology Reviews*, 8(1):14–25, jun 2012.

[21] Yuriy Kurylyak, Francesco Lamonaca, and Domenico Grimaldi. A Neural Network-based method for continuous blood pressure estimation from a PPG signal. In *Conference Record - IEEE Instrumentation and Measurement Technology Conference*, pages 280–283, 2013.

[22] Aman Gaurav, Maram Maheedhar, Vijay N. Tiwari, and Rangavittal Narayanan. Cuff-less PPG based continuous blood pressure monitoring - A smartphone based approach. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, volume 2016-Octob, pages 607–610. Institute of Electrical and Electronics Engineers Inc., oct 2016.

[23] Mustafa Radha, Koen De Groot, Nikita Rajani, Cybele Cp Wong, Nadja Kobold, Valentina Vos, Pedro Fonseca, Nikolaos Mastellos, Petra A Wark, Nathalie Velthoven, Reinder Haakma, and Ronald M Aarts. Estimating blood pressure trends and the nocturnal dip from photoplethysmography. Technical report, 2018.

[24] S. H. Liu, S. H. Lai, J. J. Wang, T. H. Tan, and Y. F. Huang. The Cuffless Blood Pressure Measurement with Multi-dimension Regression Model based on Characteristics of Pulse Waveform. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 6838–6841. Institute of Electrical and Electronics Engineers Inc., jul 2019.

[25] Satya Narayan Shukla. Estimation of blood pressure from non-invasive data. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pages 1772–1775. Institute of Electrical and Electronics Engineers Inc., sep 2017.

[26] Ludi Wang, Wei Zhou, Ying Xing, and Xiaoguang Zhou. A Novel Neural Network Model for Blood Pressure Estimation Using Photoplethesmography without Electrocardiogram. *Journal of healthcare engineering*, 2018:7804243, 2018.

[27] Mohammad Reza Mohebbian, Anh Dinh, Khan Wahid, and Mohammad Sami Alam. Blind, Cuff-less, Calibration-Free and Continuous Blood Pressure Estimation using Optimized Inductive Group Method of Data Handling. *Biomedical Signal Processing and Control*, 57:101682, mar 2020.

[28] Francesco Rundo, Alessandro Ortis, Sebastiano Battiato, and Sabrina Conoci. Advanced Bio-Inspired System for Noninvasive Cuff-Less Blood Pressure Estimation from Physiological Signal Analysis. *Computation*, 6(3):46, aug 2018.

[29] Shi Chao Gao, Peter Wittek, Li Zhao, and Wen Jun Jiang. Data-driven estimation of blood pressure using photoplethysmographic signals. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, volume 2016-Octob, pages 766–769. Institute of Electrical and Electronics Engineers Inc., oct 2016.

[30] Gašper Slapničar, Nejc Mlakar, and Mitja Luštrek. Blood Pressure Estimation from Photoplethysmogram Using a Spectro-Temporal Deep Neural Network. *Sensors*, 19(15):3420, aug 2019.

[31] de Carvalho. *Is Pulse Transit Time a Good Indicator of Blood Pressure Changes During Short Physical Exercise in a Young Population?* 2010.

[32] W. Chen, T. Kobayashi, S. Ichikawa, Y. Takeuchi, and T. Togawa. Continuous estimation of systolic blood pressure using the pulse arrival time and intermittent calibration. *Medical and Biological Engineering and Computing*, 38(5):569–574, 2000.

[33] Carmen C Y Poon, Yuan-Ting Zhang, and Mico AE Yee-Man Wong Carmen Chung-Yan Poon AE Yuan-Ting Zhang. An Evaluation of the Cuffless Blood Pressure Estimation Based on Pulse Transit Time Technique: a Half Year Study on Normotensive Subjects High-resolution imaging recognition and risk assessment of cardiovascular vulnerable plaque View project Multi-wavele. *Springer*, 9(1):32–38, 2014.

[34] HJ Baek, KK Kim, JS Kim, B Lee Physiological ..., and undefined 2009. Enhancing the estimation of blood pressure using pulse arrival time and two confounding factors. *iopscience.iop.org*.

[35] Thomas Wibmer, K Doering, C Kropf-Sanchen, S Rüdiger, I Blanta, K M Stoiber, W Rottbauer, C Schumann, and Thomas Wibmer. Pulse Transit Time and Blood Pressure During Cardiopulmonary Exercise Tests. *Physiol. Res*, 63:287–296, 2014.

[36] Younhee Choi, Qiao Zhang, and Seokbum Ko. Noninvasive cuffless blood pressure estimation using pulse transit time and Hilbert-Huang transform. *Computers & Electrical Engineering*, 39(1):103–111, jan 2013.

[37] P Fung, G Dumont, C Ries, C Mott The 26th annual ..., and undefined 2004. Continuous noninvasive blood pressure measurement by pulse transit time. *ieeexplore.ieee.org*.

[38] Michela Masè, Walter Mattei, Roberta Cucino, Luca Faes, and Giandomenico Nollo. Feasibility of cuff-free measurement of systolic and diastolic arterial blood pressure. *Journal of Electrocardiology*, 44(2):201–207, mar 2011.

[39] Heather Ting Ma. A Blood Pressure Monitoring Method for Stroke Management. *hindawi.com*, 2014.

[40] Mohamad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In *Proceedings - IEEE International Symposium on Circuits and Systems*, volume 2015-July, pages 1006–1009. Institute of Electrical and Electronics Engineers Inc., jul 2015.

[41] Mohammad Kachuee, Mohammad Mahdi Kiani, Hoda Mohammadzade, and Mahdi Shabany. Cuffless Blood Pressure Estimation Algorithms for Continuous Health-Care Monitoring. *IEEE Transactions on Biomedical Engineering*, 64(4):859–869, apr 2017.

[42] Yi Kim Jung, Hwan Cho Baek, Mi Im Soo, Ju Jeon Myoung, Young Kim In, and I. Kim Sun. Comparative study on artificial neural network with multiple

regressions for continuous estimation of blood pressure. In *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings*, volume 7 VOLS, pages 6942–6945, 2005.

[43] Peng Su, Xiao Rong Ding, Yuan Ting Zhang, Jing Liu, Fen Miao, and Ni Zhao. Long-term blood pressure prediction with deep recurrent neural networks. In *2018 IEEE EMBS International Conference on Biomedical and Health Informatics, BHI 2018*, volume 2018-Janua, pages 323–328. Institute of Electrical and Electronics Engineers Inc., apr 2018.

[44] Solmaz Rastegar, Hamid Gholamhosseini, Andrew Lowe, Farhad Mehdipour, and Maria Lindén. Estimating Systolic Blood Pressure Using Convolutional Neural Networks. *Studies in health technology and informatics*, 261:143–149, 2019.

[45] Sanghyun Baek, Jiyong Jang, and Sungroh Yoon. End-to-end blood pressure prediction via fully convolutional networks. *IEEE Access*, 7:185458–185468, 2019.

[46] Abbas K. Abbas and Rasha Bassam. Phonocardiography Signal Processing. *Synthesis Lectures on Biomedical Engineering*, 31:1–189, apr 2009.

[47] Kyoung Hoon Lim, Young Duck Shin, Sang Hi Park, Jin Ho Bae, Hong Jae Lee, Seon Jung Kim, Ji Yun Shin, and Young Jin Choi. Correlation of blood pressure and the ratio of S1 to S2 as measured by esophageal stethoscope and wireless bluetooth transmission. *Pakistan Journal of Medical Sciences*, 29(4):1023, 2013.

[48] Ana Castro, Sandra S. Mattos, and Miguel T. Coimbra. Noninvasive blood pressure and the second heart sound analysis. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2014*, pages 5494–5497. Institute of Electrical and Electronics Engineers Inc., nov 2014.

[49] Tien-En Chen, Shih-I Yang, Li-Ting Ho, Kun-Hsi Tsai, Yu-Hsuan Chen, Yun-Fan Chang, Ying-Hui Lai, Syu-Siang Wang, Yu Tsao, and Chau-Chung Wu. S1 and S2 Heart Sound Recognition Using Deep Neural Networks. *IEEE transactions on bio-medical engineering*, 64(2):372–380, 2017.

[50] Gaurav Kapur, Lingguang Chen, Yong Xu, Katherine Cashen, Jeff Clark, Xiaoce Feng, and Sean F. Wu. Noninvasive Determination of Blood Pressure by Heart Sound Analysis Compared With Intra-Arterial Monitoring in Critically Ill Children-A Pilot Study of a Novel Approach. *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*, 20(9):809–816, sep 2019.

[51] Gima S.p.A. Gima ABPM pulse rate monitor - User Manual.

[52] Praveen Kumar Sharma, Sourav Saha, and Saraswati Kumari. Study and Design of a Shannon-Energy-Envelope based Phonocardiogram Peak Spacing Analysis for Estimating Arrhythmic Heart-Beat. *International Journal of Scientific and Research Publications*, 4(9), 2014.

[53] Yoshua Bengio, Ian Goodfellow, and Aaron Courville. Deep Learning. Technical report, 2015.

[54] Connor Shorten and Taghi M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, 6(1):1–48, dec 2019.

[55] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016-Decem:779–788, jun 2015.

[56] Joseph Redmon and Ali Farhadi. YOLOv3: An Incremental Improvement. apr 2018.

[57] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4):541–551, dec 1989.

[58] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. International Conference on Learning Representations, ICLR, dec 2015.