

Human Identification attraverso analisi dei segnali EEG

Mattia d'Argenio, Simone Masullo

m.dargenio5@studenti.unisa.it, s.masullo7@studenti.unisa.it

Abstract

L'identificazione di un soggetto attraverso dati biometrici sta acquisendo sempre maggiore favore sia per motivi forensi che per motivi di attribuzione di identità all'interno di un sistema informatico. L'identificazione di un soggetto umano viene comunemente effettuata tramite riconoscimento dell'impronta digitale, del volto o dell'iride, ma esistono molte altre biometrie che offrono diversi vantaggi e svantaggi. Tra queste vi sono i segnali elettro encefalografici (Electroencephalogram - EEG), ovvero segnali elettrici provenienti dal cervello. Il seguente articolo ha lo scopo di mostrare un approccio all'utilizzo dei segnali EEG per creare e risolvere un'istanza del problema comune in computer vision conosciuto sotto il nome di Human Identification.

1 Introduction

L'identificazione dell'individuo, o human identification, è un comune problema che coinvolge tecniche di machine learning o deep learning volte al riconoscimento di un soggetto. Solitamente il processo di identificazione biometrica prevede tre fasi. La prima è l'acquisizione del fattore biometrico attraverso appositi sensori. La seconda è il preprocessing dell'informazione ottenuta per ottenere dati rilevanti e spendibili nell'ambito della classificazione del soggetto. Infine viene effettuata la classificazione vera e propria, che prevede l'utilizzo di un modello pre-addestrato in grado di riconoscere il soggetto in questione. Il seguente articolo illustra la realizzazione di un progetto in ambito universitario per l'utilizzo dei segnali elettroencefalografici come fattore per l'identificazione biometrica. L'articolo è strutturato come di seguito. Nella Sezione 2 è presente

una breve introduzione ai segnali EEG per essere in grado di dare un contesto ai dati studiati. Nella Sezione 3 viene fatta una descrizione del dataset utilizzato per il testing del progetto. Nella Sezione 4 vengono illustrate le tecniche di preprocessing utilizzate in letteratura per i sistemi di identificazione biometrica basata sui segnali EEG. Nella Sezione 5 vengono mostrati i risultati delle tecniche di Machine Learning o Deep Learning utilizzate in letteratura. Nella Sezione 6 vengono mostrata una soluzione sperimentale e ne vengono discussi i risultati. Nella Sezione 7 sono presenti le conclusioni dell'articolo.

2 I segnali EEG

Un segnale EEG è ottenuto dalla registrazione dell'attività elettrica dei neuroni. Al fine di acquisire tali segnali vengono posizionati degli elettrodi in punti specifici della testa del soggetto. Quando si analizzano questi segnali bisogna tener conto della distinzione degli stessi in cinque gruppi, ottenuti al variare delle frequenze:

- **Delta 0.5-4 Hz:** sono i segnali generati dal cervello durante fasi di sonno profondo;
- **Theta 4-8 Hz:** sono i segnali generati durante fasi di concentrazione o meditazione;
- **Alpha 8-13 Hz:** sono i segnali generati dal cervello durante fasi di rilassamento;
- **Beta 14-25Hz:** sono i segnali generati durante fasi di ragionamento;
- **Gamma >25 Hz:** sono i segnali generati durante l'elaborazione di informazioni più o meno complesse.

2.1 EEG come fattore per l'identificazione biometrica

Oggigiorno il processo di identificazione di un soggetto viene effettuato tramite l'utilizzo di

biometrie che sono veloci da acquisire e che possono rappresentare un fattore discriminante, ovvero che varia da soggetto a soggetto. Tra le biometrie più famose al fine di ottenere questo obiettivo troviamo le impronte digitali, il volto e l'iride. Le tre biometrie appena menzionate hanno il vantaggio di essere facili da acquisire, con attrezzature che ad oggi risultano essere poco costose. Lo stesso non si può dire dei segnali EEG, per l'acquisizione dei quali non esistono attrezzature davvero economiche. Le sessioni di acquisizione al fine di identificazione possono essere lunghe e richiedere ambienti preparati ad hoc. Nonostante ciò, i segnali EEG offrono diversi vantaggi nel campo dell'identificazione o autenticazione di un soggetto, primo tra tutti la difficoltà di falsificazione. Al fine di registrare segnali derivanti dall'attività del cervello, innanzitutto il cervello deve essere in attività. Ciò significa che, calandosi nel contesto dell'autenticazione a sistemi ad alta sicurezza, non è possibile asportare questo fattore biometrico da un soggetto. Se ciò non bastasse, affinché i segnali risultino inalterati, non è possibile sottoporre un soggetto a condizioni di stress, dunque egli non potrebbe essere costretto ad effettuare l'autenticazione ad un sistema.

3 Dataset Utilizzato

Nell'ambito del seguente studio è stato prodotto un progetto volto a testare tecniche di preprocessing e algoritmi di machine learning e deep learning per la classificazione dei segnali EEG e per l'identificazione di diversi soggetti. Al fine di avere una base di dati strutturata e documentata per l'esecuzione dei test menzionati, è stato scelto Biometric EEG Dataset (BED) [(Arnau-González et al., 2021)]. La documentazione fornita a corredo di questo dataset consente di inquadrare e valutare la bontà delle acquisizioni effettuate. Il dataset fornisce acquisizioni per 21 soggetti per ognuno dei quali vengono effettuate tre sessioni di acquisizione, dilazionate nel tempo.

3.1 Struttura del dataset

Così come documentato dagli autori del dataset BED, i file forniti sono suddivisi in tre categorie:

- **RAW:** contiene le misurazioni in formato csv di 21 soggetti, per ognuno dei quali sono disponibili 3 sessioni. Per i dati raw non è stato effettuato alcun tipo di preprocessing o di feature extraction;

- **RAW PARSED:** contiene 63 file in formato tabelle matlab, riconducibili al soggetto ed alla sessione di acquisizione relativa grazie al nome assegnatogli. I nomi dei file sono tutti formattati come "sX_sY.mat" dove X indica l'identificativo del soggetto e Y il numero della sessione. Nel corso del progetto sviluppato i soggetti saranno identificati da tale valore numerico X. I valori delle misurazioni presenti nei file menzionati non sono stati manipolati attraverso preprocessing;
- **Features:** comprende le features estratte a partire dai dati delle misurazioni. Tali features sono estratte tramite tre tecniche, AutoRegression Reflection Coefficients (ARRC), Mel Frequency Cepstral Coefficients (MFCC) e Spectral Features (SPEC). Ulteriori informazioni riguardanti le tecniche di preprocessing e di feature extraction adoperate dagli autori del dataset BED sono fornite nell'articolo pubblicato di riferimento.

Tra i dati disponibili si è scelto di utilizzare quelli appartenenti alla categoria Raw Parsed. L'assenza di tecniche di preprocessing ha permesso di lavorare con dati non contaminati.

4 Tecniche di preprocessing dei dati

Al fine di comprendere quanto presente nella seguente sezione occorre riassumere brevemente lo scopo del preprocessing dei dati fornendo la terminologia adeguata. Durante l'acquisizione di segnali, nel caso particolare di segnali provenienti dai neuroni, possono essere presenti fattori che comportano la contaminazione dei dati ottenuti. In generale, quando si trattano dati ai fini biometrici, è importante applicare tecniche di preprocessing al fine di eliminare tutto ciò che è considerato rumore. L'eliminazione del rumore e l'attenuazione degli errori dovuti all'acquisizione possono essere suddivise in quattro categorie:

- **rimozione dei canali sporchi:** durante un'acquisizione di segnali EEG si ottengono misurazioni provenienti da diverse fonti, dette canali. Ogni canale viaggia su diverse frequenze. Accade talvolta che le misurazioni presentino valori palesemente alterati. Ciò è dovuto al malfunzionamento di elettrodi o al cattivo posizionamento degli

stessi. Per evitare di lavorare su dati errati, è necessario rimuovere questi canali;

- **interpolazione:** per interpolazione si intende la predizione di un valore basata sui valori che si hanno a disposizione. Una volta identificato e rimosso un canale con dati errati, vorremmo avere comunque a disposizione dei dati corretti per eseguire successivamente delle analisi. È possibile predire i valori corretti di un canale errato a partire dai dati degli altri canali;
- **filtering:** le misurazioni ottenute durante la fase di acquisizione dei segnali non sono sempre utili, dunque alcune vanno rimosse. Ci si riferisce in questi casi a misurazioni non errate, ma di elementi non appartenenti al dominio di riferimento. Ricadono in questa categoria, ad esempio, le frequenze alte che vengono rilevate durante un'acquisizione che non necessariamente corrispondono ad un'errata misurazione;
- **artifact rejection:** similmente a quanto accade per il filtraggio, spesso è necessario rimuovere le acquisizioni provenienti da eventi o elementi che vanno ignorati poiché potrebbero distogliere l'attenzione da ciò che è realmente importante. Nel contesto dei segnali EEG, ad esempio, rientrano in questa categoria le misurazioni del segnale elettrico. Esso proviene dagli elettrodi utilizzati per l'acquisizione, e presenta sempre un valore di 50 Hz, ma non fa parte dei segnali encefalografici dunque va ignorato.

4.1 Preprocessing eseguito

Come menzionato in (Abo-Zahhad et al., 2015) la tecnica più comune per il preprocessing dei segnali EEG al fine del riconoscimento di esseri umani è l'applicazione di un filtro passa banda che limiti l'analisi a valori compresi tra 0.2 Hz e 50 Hz. Tali valori, presi in analisi in (Ravi and Palaniappan, 2005), hanno lo scopo di porre un limite inferiore ai valori presi in considerazione al fine di escludere valori poco significativi per l'identificazione. Il limite superiore è dato dal fatto che a 50 Hz si trovano i segnali elettrici generati dagli elettrodi utilizzati per l'acquisizione. In aggiunta ai filtri sono state identificate altre tecniche di preprocessing comunemente utilizzate nel contesto dei segnali EEG.

Tali tecniche sono riassunte in quello che è ormai diventato uno standard per il preprocessing dei segnali EEG, chiamato PREP pipeline (Bigdely-Shamlo et al., 2015). In aggiunta alle tecniche di filtering, nel corso del progetto sviluppato sono stati effettuati diversi tentativi di preprocessing, tra cui Independent Component Analysis e Principal Component Analysis al fine di ridurre la dimensionalità dei dati andando ad estrarre unicamente le misurazioni più significative. È importante notare che il dataset su cui si basa il progetto svolto prevede una frequenza di campionamento di 256 Hz/s. Una frequenza di campionamento così alta può comportare la ripetizione di valori molto simili tra loro, dunque un'estrazione di componenti indipendenti è in particolar modo efficiente in tale situazione.

4.2 Feature extraction

Durante l'analisi di segnali è buona norma estrarre delle features che possano essere più significative delle misurazioni stesse. La fase della feature extraction è utilizzata per trasformare dati raw in valori numerici che rappresentino aspetti dei dati dal punto di vista statistico. Come menzionato nella sezione 3 relativa al Dataset BED, nel paper fornito a corredo del dataset vengono già introdotte delle tecniche di feature extraction utilizzate, ed i cui risultati sono reperibili. Tali tecniche sono: computazione di Auto Regression Reflection Coefficients, di Mel Frequency Cepstral Coefficients e di Spectral Features.

4.2.1 Auto Regression Reflection Coefficients

I coefficienti ARRC sono dei coefficienti estratti a partire dal modello di Auto Regression (AR model) utilizzando l'equazione di Yule-Walker. Le features estratte tramite questa tecnica sono valori puramente matematici che indicano la variazione del segnale attorno ad una media computata a partire dai valori stessi. Una tecnica simile viene presentata da (Ashby et al., 2011) per risolvere il problema dell'autenticazione attraverso segnali EEG.

4.2.2 Mel Frequency Cepstral Coefficients

Le tecniche di feature extraction basate su MFCC sono state comunemente utilizzate nel contesto dei segnali sonori, ad esempio per la speech recognition. Tuttavia sono state recentemente mostrate diverse similitudini tra la distribuzione dei valori in un segnale audio e quella presente in una ses-

sione di acquisizione di segnali EEG (Maiorana and Campisi, 2018).

4.2.3 Spectral Features

È tecnica comune rappresentare la potenza di un segnale come funzione delle sue frequenze. Per ottenere ciò spesso si utilizza il dominio spettrale, ottenibile calcolando la Power Spectral Density di un segnale. All'interno di questo dominio è possibile computare diversi valori statistici che possono essere più o meno significativi per l'obiettivo che si intende raggiungere. Le misurazioni statistiche sono state utilizzate in altri lavori per la risoluzione dei problemi di speech recognition (Monge-Álvarez et al., 2019). A seguito delle considerazioni fatte in precedenza sulle similitudini analisi dei segnali audio ed analisi dei segnali EEG, è stata presa in considerazione la possibilità di considerare features estratte da questo dominio. Al fine di comprendere al meglio le seguenti formule si anticipa che:

- $PSD_j(k)$ rappresenta il valore del segnale k all'interno del dominio spettrale. Tale dominio è calcolato utilizzando la tecnica di Welch (Rooney and Buck, 2019). Il pedice j indica la banda di frequenza del segnale;
- $f_j(k)$ rappresenta il vettore delle frequenze discrete associate al segnale k .

Si procede ora ad illustrare i valori che si ritiene siano più consoni per l'obiettivo preposto:

- **Spectral centroid:** il centroide in statistica è il valore più rappresentativo dell'insieme dei valori, dato dalla media pesata delle frequenze.

$$\frac{\sum_k f_j(k) \times PSD_j(k)}{\sum_k PSD_j(k)}$$

- **Spectral bandwidth:** rappresenta l'ampiezza dello spettro delle frequenze in termini di valori assunti, data dalla varianza ponderata dei valori rispetto al centroide. Più bassa è l'ampiezza, più i valori delle frequenze si concentrano attorno al centroide.

$$\frac{\sum_k (f_j(k) - SpecCent(j))^2 \times PSD_j(k)}{\sum_k PSD_j(k)}$$

- **Spectral crest factor:** calcola la presenza di picchi nello spettro delle frequenze, data dal rapporto del valore più alto presente in una banda di frequenza e la distanza tra il valore

più alto e quello più basso presenti nel vettore delle frequenze discrete. Un valore alto indica la presenza di picchi alti.

$$C = \frac{1}{\max\{f_j(k)\} - \min\{f_j(k)\} + 1}$$

$$SpecCrestFac(j) = \frac{\max\{PSD_j(k)\}}{C \times \sum_k PSD_j(k)}$$

- **Spectral flatness:** indica l'omogeneità dei valori, data dal rapporto tra la media geometrica dei valori e la media aritmetica. Più il valore è alto, più i valori dello spettro delle frequenze sono ripetuti.

$$SpecFlatness(j) = \frac{\exp(\prod_k \log(PSD_k(k)))}{\prod_k PSD_j(k)}$$

4.2.4 Wavelet Transform

Così come accade per il Power Spectral Density, è possibile rappresentare un segnale in differenti domini così da poter estrarre valori statistici che andranno a comporre le features. La trasformata di Wavelet consente di rappresentare un segnale mediante l'uso di onde oscillanti di lunghezza finita. Come fatto in altri lavori (Shoka et al., 2019), le features estratte dal dominio di Wavelet sono:

- **Entropia**
- **Deviazione standard**
- **Varianza**

5 Classificazione

Al fine di riconoscere un individuo sulla base dei propri dati biometrici, sia che si desideri utilizzare i dati appena acquisiti o le features estratte da tali dati, è necessario adoperare tecniche di machine learning o deep learning. Entrambe le categorie di soluzioni fanno uso di intelligenza artificiale per addestrare modelli su un insieme di dati e per utilizzare i modelli addestrati al fine di generare predizioni sulla base dei valori forniti in input. Vengono dunque identificate due fasi fondamentali nella costruzione di un modello che sfrutta l'intelligenza artificiale, l'addestramento (training) e la valutazione del modello (testing). Quando l'insieme dei dati non prevede una divisione in training e testing, si possono adottare diverse soluzioni. La più comune è quella di dividere casualmente il dataset in una porzione per il training e, la restante parte, per il testing. I modelli di divisione prendono spesso il nome dalle percentuali di dati utilizzate per la divisione, ad esempio è comune adottare una soluzione 70:30.

5.1 Lavori correlati

Come menzionato in (Jalaly Bidgoly et al., 2020) diverse tecniche di machine learning o deep learning sono state utilizzate in letteratura per l'identificazione umana attraverso segnali EEG. (Bashar et al., 2016) fa uso di Support Vector Machine (SVM) basandosi su features estratte meglio documentate all'interno del paper, raggiungendo un'accuracy del 94,44%. Altri lavori preferiscono creare un algoritmo ad hoc per la classificazione. Nello specifico, in (Huang et al., 2012) viene costruita una rete neurale che riesce a raggiungere il 95,1% di accuracy. In (Abdulrahman et al., 2020) viene effettuato il paragone tra una SVM - già menzionata precedentemente - ed una soluzione che usa K-Nearest Neighbors(KNN) ottenendo un risultato del 99,1% attraverso l'utilizzo di una SVM che riceve dati estratti da Wavelets. In altri lavori, invece, vengono adottate soluzioni che prevedono l'uso di algoritmi che si basano sul deep learning. Tra questi, (Tabar and Halici, 2016) utilizza una rete neurale convoluzionale (CNN) ottenendo una accuracy del 90%.

6 Soluzione proposta

La premessa su cui si basa questa sezione è che la soluzione presentata è frutto di diversi tentativi nati allo scopo di acquisire familiarità con le diverse tecniche di preprocessing, feature extraction e classificazione presentate nei lavori correlati. La soluzione presentata è divisa in tre moduli principali, essi sono:

- **loading.py** all'interno del quale vengono presentate funzioni per la lettura del dataset ed il caricamento dei dati in esso presenti;
- **preprocessing.py** all'interno del quale vengono presentate funzioni per eseguire il preprocessing e la feature extraction dei dati forniti dal dataset;
- **classification.py** all'interno del quale vengono esplorate diverse possibili soluzioni per la creazione, l'addestramento ed il testing di un modello basato sull'intelligenza artificiale, oltre a fornire criteri per la divisione dei dati e per la valutazione dei risultati raggiunti.

6.1 Loading

Il caricamento dei dati potrebbe sembrare un problema banale ma si ritiene opportuno esplorare le

diverse soluzioni esistenti e spiegare le ragioni che hanno portato all'utilizzo di una soluzione personalizzata. I dati utilizzati per la soluzione che viene presentata, come già annunciato nella sezione 3.1 del presente articolo, ricadono nella categoria RAW PARSED. Durante il caricamento dei dati in memoria ci si è posto il problema di quale struttura dati utilizzare. Esistono diversi moduli in Python che promettono di fornire soluzioni integrate per la manipolazione e l'utilizzo dei segnali EEG. Tra questi, il più famoso è PyEEG (Bao et al., 2011). PyEEG fornisce una struttura dati, chiamata RAW, che permette il caricamento dei dati da un dataset e la manipolazione degli stessi. Purtroppo si tratta di una struttura che occupa molta memoria e che richiede molta conoscenza della libreria per essere utilizzata. Un'alternativa appetibile è la libreria pandas per Python, che realizza delle strutture dati chiamate Dataframe del tutto riconducibili a tabelle. Seppur sia vero che pandas offre funzioni per la manipolazione dei dati efficienti, la conversione di una grande mole di dati - come quella che è stata fornita dal dataset - in un dataframe pandas richiede molto tempo, rallentando l'esecuzione del software senza fornire un apporto particolarmente migliorativo. Ciò detto, si è proceduto a realizzare una classe Measurement la quale memorizza i nomi dei sensori presenti nel dataset, i valori delle misurazioni degli stessi, l'id del soggetto a cui le misurazioni si riferiscono ed altre informazioni utili per la realizzazione della soluzione che viene presentata.

6.2 Preprocessing

Come anticipato nella sezione 4 dedicata al preprocessing e alla feature extraction, all'interno del modulo preprocessing è possibile trovare diverse funzioni che agiscono sui dati di oggetti di tipo Measurement ed effettuano le trasformazioni necessarie. Tra queste funzioni troviamo un filtro passa-banda applicato al dominio delle frequenze che ha lo scopo di tagliare le frequenze non significative ai fini di human identification. Sono inoltre proposte due funzioni: Independent Component Analysis (ICA) e Principal Component Analysis (PCA) che realizzano una riduzione della quantità dei dati presenti. Tra le due possibilità si è osservato che l'algoritmo PCA è computazionalmente più efficiente oltre ad essere più flessibile sulla scelta della dimensione dell'insieme output. Infine sono presenti funzioni per l'estrazione di

features, ovvero PSD e Wavelet.

6.3 Classification

Ad ogni soggetto del dataset è stata attribuita una etichetta che va da 1 a 21. I modelli addestrati ed utilizzati sono tutti stati generati a partire da algoritmi di Deep Learning. La scelta è ricaduta su tale branca del Machine Learning poiché sono stati riscontrati diversi problemi nell'utilizzo delle features estratte ai fini della classificazione. Infatti, grazie agli algoritmi di deep learning, è stato possibile utilizzare i dati così come presenti nel dataset ottenendo delle metriche soddisfacenti. Gli algoritmi utilizzati sono stati: Long Short-Term Memory, Random Forest, Random Forest con grid search per la computazione dei parametri migliori per la classificazione e Random Forest attraverso la libreria XGBoost per un miglioramento delle performance. Oltre agli algoritmi per la creazione di modelli, in questo modulo vengono presentate due alternative per la divisione del dataset tra training data e test data:

- **Random split** realizza una divisione casuale dei dati forniti dal dataset in relazione ad una percentuale parametrizzata. Il valore di default del parametro in questione è 70, stabilendo dunque che il 70% dei dati viene utilizzato per l'addestramento del modello ed il restante 30% viene utilizzato per effettuare le predizioni - in caso di assenza di parametri che definiscano diversamente;
- **Session split** siccome le misurazioni della stessa sessione presentano valori molto simili tra di loro - come già fatto notare in precedenza la frequenza di acquisizione dei dati è di 256 Hz/sec comportando la presenza di valori ripetuti - si è proceduto a realizzare una divisione del dataset che prevedesse l'utilizzo delle misurazioni presenti in due sessioni di acquisizione come dati di addestramento, ed i restanti dati - dunque quelli relativi all'ultima sessione di acquisizione - come dati di testing.

In aggiunta a questi metodi, il modulo fornisce due funzioni allo scopo di valutare il modello utilizzato. La prima funzione calcola la matrice di confusione a partire da un insieme di coppie dove il primo elemento è la predizione effettuata dal modello ed il secondo è la predizione attesa. La seconda funzione calcola l'accuratezza, la precisione,

la recall e lo f1-score in un sistema multiclasse come quello presentato nel presente lavoro.

6.4 Risultati raggiunti

In questa sezione vengono presentati i risultati raggiunti tramite le metodologie elencate precedentemente. Nella tabella 1 di seguito vengono mostrate le metriche calcolate per ogni algoritmo eseguito con i due diversi tipi di split. La tabella mostra come sia stato possibile raggiungere un'accuratezza decisamente alta utilizzando l'algoritmo Random Forest. Il test sul secondo tipo di split è particolarmente necessario poiché, raggiungendo un'accuratezza alta anche in questi casi, si preclude l'ipotesi di un overfitting.

Classificatore	Tecnica di split	Accuratezza	Precisione	Recall	F1-score
LSTM	Random split	20,15%	-N/A-	-N/A-	-N/A-
LSTM	Session split	19,47%	-N/A-	-N/A-	-N/A-
Random Forest	Random split	99,98%	99,99%	99,99%	0,99
Random Forest	Session split	99,99%	99,99%	94,10%	0,96
Random Forest + Grid Search	Random split	99,99%	99,99%	99,99%	0,99
Random Forest + Grid Search	Session split	98,84%	97,70%	95,12%	0,97
Random Forest (XGBoost)	Random split	94,93%	94,93%	94,98%	0,97
Random Forest (XGBoost)	Session split	91,82%	80,91%	83,21%	0,60

Table 1: risultati delle esecuzioni.

7 Conclusioni

Il lavoro svolto è servito a mostrare le tecniche di preprocessing più comuni che riguardano i segnali EEG. Date le premesse si ritiene possibile applicare tecniche simili anche ad altri segnali elettrici. È stato particolarmente interessante osservare come diversi domini di rappresentazione dei dati possono concorrere alla produzione di features spendibili in fase di classificazione. Sebbene sia stato interessante osservare il preprocessing dei dati è stato mostrato come algoritmi di Deep Learning siano in grado di effettuare previsioni soddisfacenti malgrado l'utilizzo di dati provenienti dalle misurazioni e non frutto di accurata elaborazione. Per quanto questo rappresenti una semplificazione dal punto di vista progettuale, l'approccio basato su deep learning richiede tempi di computazione elevati, talvolta inaccettabili in alcune applicazioni.

References

- Shaymaaadan Abdulrahman, Mohamed Roushdy, and Abdel-Badeeh M Salem. 2020. Support vector machine approach for human identification based on eeg signals. *Journal of Mechanics of Continua and Mathematical Sciences*, 15(2):270–280.
- Mohammed Abo-Zahhad, Sabah Mohammed Ahmed, and Sherif Nagib Abbas. 2015. State-of-the-art methods and future perspectives for personal recognition based on electroencephalogram signals. *IET Biometrics*, 4(3):179–190.
- Pablo Arnau-González, Stamos Katsigiannis, Miguel Arevalillo-Herráez, and Naeem Ramzan. 2021. Bed: A new data set for eeg-based biometrics. *IEEE Internet of Things Journal*, 8(15):12219–12230.
- Corey Ashby, Amit Bhatia, Francesco Tenore, and Jacob Vogelstein. 2011. Low-cost electroencephalogram (eeg) based authentication. In *2011 5th International IEEE/EMBS Conference on Neural Engineering*, pages 442–445.
- Forrest Sheng Bao, Xin Liu, Christina Zhang, et al. 2011. Pyeeg: an open source python module for eeg/meg feature extraction. *Computational intelligence and neuroscience*, 2011.
- Md. Khayrul Bashar, Ishio Chiaki, and Hiroaki Yoshida. 2016. Human identification from brain eeg signals using advanced machine learning method eeg-based biometrics. In *2016 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES)*, pages 475–479.
- Nima Bigdely-Shamlo, Tim Mullen, Christian Kothe, Kyung-Min Su, and Kay A. Robbins. 2015. The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in Neuroinformatics*, 9.
- Xu Huang, Salahiddin Altahat, Dat Tran, and Dharmendra Sharma. 2012. Human identification with electroencephalogram (eeg) signal processing. In *2012 International Symposium on Communications and Information Technologies (ISCIT)*, pages 1021–1026.
- Amir Jalaly Bidgoly, Hamed Jalaly Bidgoly, and Zeynab Arezoumand. 2020. A survey on methods and challenges in eeg based authentication. *Computers and Security*, 93:101788.
- Emanuele Maiorana and Patrizio Campisi. 2018. Longitudinal evaluation of eeg-based biometric recognition. *IEEE Transactions on Information Forensics and Security*, 13(5):1123–1138.
- Jesús Monge-Álvarez, Carlos Hoyos-Barceló, Luis Miguel San-José-Revuelta, and Pablo Casaseca-de-la Higuera. 2019. A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features. *IEEE Transactions on Biomedical Engineering*, 66(8):2319–2330.
- K.V.R. Ravi and R. Palaniappan. 2005. Leave-one-out authentication of persons using 40 hz eeg oscillations. In *EUROCON 2005 - The International Conference on "Computer as a Tool"*, volume 2, pages 1386–1389.
- Ian M. Rooney and John R. Buck. 2019. Spatial power spectral density estimation using a Welch coprime sensor array processor. *The Journal of the Acoustical Society of America*, 145(4):2350–2362, 04.
- Athar Shoka, Mohamed Dessouky, Ahmed el sherbeny, and Ayman El-Sayed. 2019. Literature review on eeg preprocessing, feature extraction, and classifications techniques. *Menoufia Journal of Electronic Engineering Research*, 28:292–299, 12.
- Yousef Rezaei Tabar and Ugur Halici. 2016. A novel deep learning approach for classification of eeg motor imagery signals. *Journal of neural engineering*, 14(1):016003.