Emotion GAIT Knn Dynamic Time Warping

Samuele Antonio Cesaro, Daniele Gregori

Università degli Studi di Salerno, Via Giovanni Paolo II 132, Fisciano (SA) - 84084, Italia

ARTICLE INFO

Keywords:
Emotion Gait
Yolov7
Knn
Dynamic Time Warping
Pose Estimation
Machine Learning
Sequentia
Feature Extraction

ABSTRACT

Il riconoscimento delle emozioni di un individuo è un settore di ricerca in continua evoluzione di fondamentale importanza in diversi campi, che spaziano dalla sicurezza ai videogiochi. Un sottodominio meno esplorato riguarda il modo in cui le emozioni influenzano l'andatura, ovvero lo stile di camminata di una persona. Questo lavoro mira ad analizzare il concetto di "emotion gait" basandosi sui correlati fisici dell'espressione emotiva osservati durante la camminata. Sono stati utilizzati dati provenienti da una serie di video etichettati che mostrano individui che si avvicinano alla telecamera e comunicano una delle quattro emozioni: felicità, rabbia, tristezza e neutralità attraverso il loro modo di camminare. Le pose relative sono state estratte al fine di identificare l'emozione espressa nel video. L'approccio proposto prevede l'estrazione manuale delle caratteristiche dai video delle camminate, seguite dalla classificazione utilizzando un classificatore K-Nearest Neighbors adattato per utilizzare il Dynamic Time Warping (DTW) anziché la classica distanza Euclidea. La metodologia presentata è stata valutata utilizzando il dataset "EWalk".

1. Introduzione

L'espressione delle emozioni è una caratteristica intrinseca dell'essere umano e riveste un ruolo fondamentale nelle interazioni sociali. Oltre ai segnali emotivi tradizionali come il linguaggio verbale e le espressioni facciali, comunemente riconosciuti come indicatori primari delle emozioni, il modo in cui camminiamo può fornire importanti informazioni sul nostro stato emotivo attraverso lo studio di specifici pattern motori e del linguaggio del corpo Castellano, Villalba and Camurri (2007); Wallbott (1998).

Negli ultimi anni, la ricerca nell'ambito dell'emotion gait ha suscitato un crescente interesse tra gli scienziati e gli esperti di intelligenza artificiale, poiché si è rivelato utile in diversi settori. Ad esempio, nell'ambito medico può essere applicato per il riconoscimento delle emozioni nei bambini con disturbi dello spettro autistico Jamil, Khir, Ismail and Razak (2015) o per verificare l'insorgenza della malattia di Parkinson. Nel campo della sicurezza, l'emotion gait può aiutare a riconoscere atteggiamenti sospetti in zone ad alto rischio, mentre nello sviluppo di robot emotivamente consapevoli può contribuire a migliorare l'interazione uomo-macchina.

Come precedentemente accennato, le emozioni umane possono essere dedotte attraverso lo studio di diverse biometriche, come le espressioni facciali, l'andatura, i gesti con le mani e il testo. Sebbene il riconoscimento delle emozioni sia un settore già da tempo largamente studiato, lo studio di sistemi in grado di riconoscere le emozioni attraverso l'andatura ha preso piede solo di recente. Questo interesse è attribuibile ai molteplici benefici che questo approccio offre rispetto ad altre modalità. Ad esempio, l'andatura fornisce una metodologia non invasiva che consente la raccolta di dati a distanza, anche su soggetti non cooperativi (Chiu, Shu and Hui (2018); Xu, Fang, Hu, Ngai, Wang, Guo and Leung (2022)). Infatti, gli esseri umani possiedono un'abilità innata di esprimere le emozioni attraverso i movimenti fisici. Quando proviamo gioia, ad esempio, la nostra andatura può diventare più vivace, mostrando una maggiore leggerezza nei movimenti. Al contrario, quando ci sentiamo tristi, la nostra andatura può diventare più lenta e pesante (Xu et al. (2022)). Questi sottili cambiamenti nel linguaggio del corpo sono i cosiddetti segnali non verbali, che ci consentono di trasmettere ed interpretare le emozioni inconsciamente.

L'articolo è organizzato come segue. Nella seconda sezione, presenteremo gli studi più rilevanti condotti finora per identificare e riconoscere le emozioni attraverso l'andatura. Nella sezione successiva, forniremo un'illustrazione dettagliata del dataset utilizzato nello studio. Nella descrizione del dataset, forniremo informazioni sull'origine dei dati, la raccolta e l'acquisizione dei campioni e le caratteristiche delle istanze presenti. Nella quarta sezione, discuteremo le metodologie utilizzate, come l'estrazione dei body joints tramite algoritmi di pose estimation, la scelta delle metriche utilizzate per la creazione delle caratteristiche, il processing applicato alle caratteristiche ottenute e il tipo di classificatore utilizzato per la sperimentazione. Infine, esporremo le nostre discussioni e conclusioni e i risultati ottenuti mediante la visualizzazione di grafici e matrici di confusione, al fine di presentare i dati in modo chiaro e comprensibile.

2. Stato dell'Arte

Anche se il settore della emotion gait recognition è relativamente nuovo, negli ultimi anni sono stati sviluppati svariati approcci che si sono rivelati interessanti.

Partendo dalle modalità di acquisizione dei dati, i ricercatori spesso utilizzano sistemi di motion capture, sensori indossabili o framework per l'estrazione delle pose (pose estimation) al fine di acquisire i movimenti degli individui mentre esprimono vari stati emotivi. Per l'estrazione delle pose, si distinguono principalmente due tipi di framework: quelli per l'estrazione delle pose 2D e quelli per l'estrazione delle pose 3D.

I framework per l'estrazione delle pose 2D si concentrano sulla stima delle posizioni dei punti di riferimento corporei

(body joints) in un'immagine bidimensionale. Tuttavia, potrebbero incontrare difficoltà in situazioni con condizioni ambientali non ottimali, come occlusioni o sfondi complessi. D'altra parte, i framework per l'estrazione delle pose 3D si occupano di ottenere le coordinate tridimensionali dei punti di riferimento corporei, consentendo una comprensione più dettagliata dei movimenti dell'individuo nello spazio tridimensionale. Tuttavia, l'utilizzo di questi framework richiede attrezzature aggiuntive, come multiple telecamere o sensori di profondità, che possono rappresentare una sfida aggiuntiva.

La maggior parte dei lavori fa uso di algoritmi di Machine Learning combinati con un'estrazione manuale delle caratteristiche, utilizzando i metodi precedentemente citati. Un esempio di studio che fa uso di questo approccio è quello descritto da Chiu et al. (2018). Questo sistema rappresenta un'implementazione su dispositivi mobili, supportata da un server cloud, che consente il riconoscimento delle emozioni anche in situazioni con più persone, come ad esempio in spazi pubblici. L'applicazione permette la classificazione di cinque emozioni: tristezza, felicità, rabbia, neutralità e rilassamento. Per l'estrazione delle caratteristiche, è stato utilizzato un algoritmo di pose estimation chiamato Open Pose, che consente di estrarre 18 punti chiave del corpo (body joints).

Sono stati allenati e confrontati sei modelli di classificazione: SVM, MLP, Decision Tree, Naive Bayes, Random Forest e Logistic Regression. Dai risultati ottenuti, il modello SVM ha dimostrato le migliori performance, con un'accuratezza del 62.1%. A seguire, il Random Forest ha raggiunto il 61.3% di accuratezza, seguito dal MLP con il 61.2%.

Ad oggi, il Deep Learning ha guadagnato notevole considerazione nel settore della emotion gait recognition grazie alla sua capacità di apprendere automaticamente caratteristiche complesse direttamente dai dati in input, senza che queste vengano definite manualmente dall'esperto. Un esempio di utilizzo di questa modalità è rappresentato dal lavoro di Mathivanan and Perumal (2022), in cui gli autori hanno sviluppato un algoritmo chiamato ADBNN-BWO, impiegando il dataset denominato *E-Walk*. Questo algoritmo combina le DBNN (Deep Belief Neural Networks) con il BWO (Black Widow Optimization) per classificare quattro emozioni: rabbia, felicità, tristezza e stupore. L'algoritmo è composto da quattro fasi: pre-processing, feature extraction, feature selection e classificazione. Nella fase di preprocessing viene applicato un filtro mediana per migliorare la qualità dei dati. Per la fase di feature extraction, vengono utilizzate diverse tecniche, tra cui Hu Moment, GLCM (Grey-Level Co-occurrence Matrix), F-SIFT (Fast Scale-Invariant Feature Transform) e pose estimation. ADBNN-BWO ha raggiunto un'accuratezza del 97%, superando significativamente gli altri algoritmi comparati nello studio.

Una direzione consistente adottata di recente nel settore è un approccio ibrido per l'estrazione delle caratteristiche, tramite la fusione dei dati ottenuti dall'estrazione manuale delle caratteristiche con le caratteristiche latenti acquisite tramite una rete neurale profonda, come nel lavoro di Bhatia (2022), in cui dimostra che concatenando le caratteristiche estratte manualmente all'output della LSTM è possibile migliorare la robustezza del modello. Questo è dovuto al fatto che i dati estratti tramite il Deep Learning sono altamente influenzati dal dataset utilizzato per l'allenamento della rete profonda. Con questa modalità operativa ha raggiunto una micro mean Average Precision di 0.98 e una macro mAP di 0.96.

3. Dataset

Il dataset utilizzato è una partizione del dataset originale intitolato *E-Walk* pubblicato da Randhavane, Bhattacharya, Kapsaskis, Gray, Bera and Manocha (2020). Questo dataset comprende 76 video di persone che camminano e cercano di esprimere una delle quattro emozioni: rabbia, felicità, tristezza o neutralità. In ogni video, i soggetti partono da una distanza di 7 metri e si avvicinano gradualmente alla telecamera. Per consentire il riconoscimento delle emozioni attraverso il movimento del corpo anziché dalle espressioni facciali, il volto è stato coperto. Ogni video ha una durata di 8 secondi e ripete ciclicamente il ciclo di gait.

Il dataset è suddiviso in un training set che contiene 61 video, e un test set, che contiene 15 video. Ogni set è accompagnato da un file Excel che include i punteggi assegnati dai partecipanti a ciascun video. Questi punteggi rappresentano un voto da 1 a 5, dove 1 indica un forte disaccordo e 5 indica un forte accordo, riguardo all'emozione percepita osservando il video. I punteggi sono stati utilizzati per classificare i video nelle quattro categorie di emozioni menzionate in precedenza.

Per ogni video, si è calcolata la media dei voti assegnati dai partecipanti per ciascuna emozione. Sulla base dei risultati ottenuti, visualizzabili nelle due figure seguenti (Figura 1 e Figura 2), è stata assegnata l'emozione che ha ottenuto un punteggio nettamente più alto rispetto alle altre. In caso di parità o valori simili tra più emozioni, è stata esclusa la neutralità e assegnata l'emozione con il punteggio più alto tra le rimanenti. Questa procedura ha portato a una distribuzione uniforme dei video tra le diverse classi (Tabella 1).

Table 1
Distribuzione dataset per emozione: la prima riga rappresenta il training set, la seconda il test set

Rabbia	Tristezza	Neutralità	Felicità
15	17	15	14
3	3	5	4

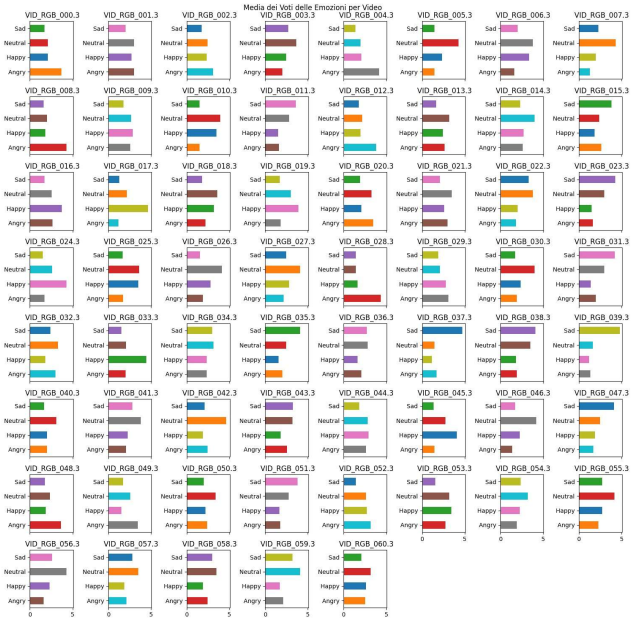

Figure 1: Medie dei voti di ciascuna emozione per ogni video presente nel training set

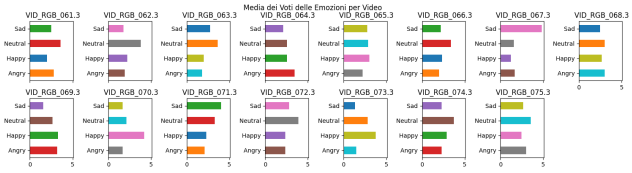

Figure 2: Medie dei voti di ciascuna emozione per ogni video presente nel test set

4. Metodologia

La seguente sezione descrive la metodologia impiegata in questo studio per l'emotion gait recognition. Essa comprende una panoramica delle tecnologie utilizzate, delle procedure di estrazione dei dati e delle tecniche di analisi dati.

4.1. Pose Estimation

I modelli di pose estimation sono algoritmi di computer vision utilizzati per ottenere informazioni sulla posa del soggetto in un'immagine o video. L'obiettivo della pose estimation è di determinare accuratamente la posizione di specifici keypoints del soggetto, i quali variano in base all'algoritmo adoperato. In questo lavoro, è stato utilizzato un modello noto in letteratura per l'object detection, e adattato per eseguire task di pose estimation, chiamato Yolov7 (Wang, Bochkovskiy and Liao (2022); Maji, Nagori, Mathew and Poddar (2022)). Tradizionalmente, i modelli di pose estimation hanno bisogno di due fasi per inferire la posa: la prima per individuare il soggetto nello spazio e la seconda per affinare l'individuazione dei keypoint. Yolo, invece, elimina la necessità di due fasi eseguendo entrambe le operazioni in un unico passaggio.
Il modello associa 17 keypoint (Figura 3 e Tabella 2) ad ogni soggetto individuato nell'immagine, e ciascun keypoint è identificato dalle corrispondenti coordinate 2D: (x, y) e il valore di confidenza(conf). La confidenza è un parametro basato sulla visibilità di un keypoint, ma tale valore viene scartato quando andiamo ad effettuare la fase di estrazione delle caratteristiche.
Poiché i soggetti nei video hanno il volto oscurato per focalizzare l'attività di riconoscimento sul corpo del soggetto, i keypoints del naso (0), degli occhi (1, 2) e delle orecchie (3, 4) non vengono presi in considerazione nella fase di estrazione delle caratteristiche.
Dato che ogni video ripete ciclicamente il ciclo di gait, la pose estimation non viene applicata all'intero video, ma solo ai frame relativi a un ciclo di gait.

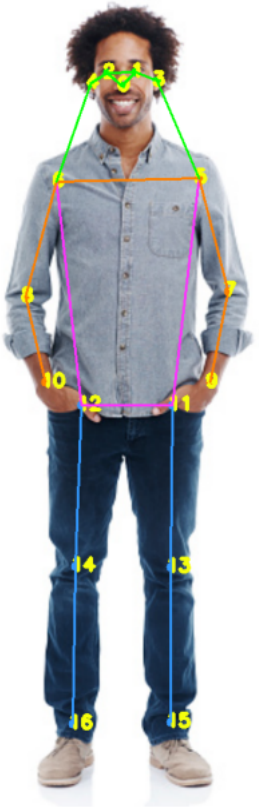

Figure 3: Keypoints generati da Yolov7

4.2. Feature Extraction

Al fine di predire accuratamente lo stato emotivo di una persona, l'analisi delle caratteristiche posturali si è dimostrata efficace, come evidenziato nel lavoro di Crenn, Khan, Meyer and Bouakaz (2016). Questi autori, a loro volta, si sono basati su studi psicologici come quello condotto da Kleinsmith and Bianchi-Berthouze (2012), in cui si afferma che: "il linguaggio del corpo è tanto potente quanto le espressioni facciali nel trasmettere le emozioni". Rappresentiamo le feature estratte da un ciclo di gait G come un vettore $F \in \mathbb{R}^{15}$. Utilizziamo un singolo ciclo di gait che corrisponde al tempo trascorso tra il primo contatto di due passi consecutivi dello stesso piede. Nel nostro framework,

Table 2
Indici dei keypoints e dei relativi joints

Indice	Descrizione
0	Naso
1	Occhio Sinistro
2	Occhio Destro
3	Orecchio Sinistro
4	Orecchio Destro
5	Spalla Sinistra
6	Spalla Destra
7	Gomito Sinistro
8	Gomito Destro
9	Mano Sinistra
10	Mano Destra
11	Bacino Sinistro
12	Bacino Destro
13	Ginocchio Sinistro
14	Ginocchio Destro
15	Piede Sinistro
16	Piede Destro

effettuiamo l'estrazione delle feature posturali per ogni fotogramma del ciclo utilizzando la rappresentazione della posa ottenuta tramite l'utilizzo di YoloV7 (Sezione 4.1). Abbiamo suddiviso le feature in tre categorie: angolo, distanza ed area. Le feature estratte sono elencate per categoria nella tabella 3. Di seguito sono spiegate le categorie considerate:

- **Distanza**: Calcolo della distanza euclidea tra due joint del corpo. Come afferma Crenn et al. (2016): "testa, bacino, gomito e spalle integrano la maggior parte dell'emozione di un movimento". Nell'articolo viene inoltre sottolineato che anche le mani rappresentano un importante mezzo espressivo. Alla luce di queste conclusioni, nello studio sono stati adoperati tutti i joints precedentemente citati, ad eccezione della testa, poiché nel nostro dataset è oscurata per focalizzarsi sul corpo del soggetto. Un esempio di distanza utilizzata è la distanza tra la mano e la spalla sullo stesso lato, che come attesta Crenn et al. (2016), riesce a convogliare indirettamente informazioni sul movimento del gomito.
- **Angolo**: Il calcolo dell'angolo ci fornisce informazioni sulla flessione delle varie articolazioni, la quale tende a cambiare in base alle diverse emozioni.Ad esempio, una flessione del gomito quasi nulla può indicare tristezza, indicando che il soggetto mantiene le braccia attaccate al corpo, mentre una flessione più accentuata può indicare felicità.
- **Area**: Un'idea introdotta inizialmente da Crenn et al. (2016), e poi ottimizzata da Randhavane et al. (2020). Mediante il calcolo dell'area dei triangoli creati durante il movimento, riusciamo a carpire infomrazioni riguardanti l'espansione del corpo durante le varie emozioni.

Table 3: **Feature posturali**: Estraiamo le feature posturali basandoci sulla letteratura psicologica

Tipo	Descrizione
Distanza	Ginocchio destro e sinistro
	Piede destro e sinistro
	Mano sinistra a spalla sinistra
	Mano destra a spalla destra
	Mano destra a mano sinistra
	Mano destra a fianco destro
	Mano sinistra a fianco sinistro
	Mano sinistra a piede sinistro
	Mano destra a piede destro
Angolo	Gomito interno sinistro (spalla, gomito, mano)
	Gomito interno destro (spalla, gomito, mano)
	Ascella sinistra (fianco, spalla, gomito)
	Ascella destra (fianco, spalla, gomito)
Area	Triangolo tra mani e punto medio spalle
	Traingolo tra piedi e punto medio fianchi

4.3. Classificazione

Le feature estratte in ogni frame formano una sequenza temporale. Quindi, ogni video sarà composto da un numero variabile di sequenze temporali, a seconda della velocità di camminata del soggetto nel video. La lunghezza variabile dei video rappresenta una sfida. Infatti, i metodi tradizionali di classificazione richiedono un campionamento delle feature per ottenere vettori di feature con una lunghezza prefissata. Tuttavia, il campionamento delle sequenze temporali può introdurre informazioni errate o eliminare parte dell'informazione, con possibili ripercussioni sulle prestazioni in fase di inferenza. Per risolvere la problematica discussa, è stato utilizzato l'algoritmo dei K-nearest Neighbours (KNN) come classificatore. Questo algoritmo fa uso del Dynamic Time Warping (DTW) come misura di distanza, anziché la tradizionale distanza euclidea.
Il DTW è un algoritmo di programmazione dinamica che tiene conto della natura temporale dei dati e consente di elaborare sequenze di lunghezza variabile. calcolando una misura di dissimilarità tra i dati di allenamento e i campioni di test senza la necessità di ricampionamento. L'algoritmo calcola l'allineamento ottimale tra due sequenze temporali costruendo dinamicamente un percorso che minimizza la distanza accumulata tra i punti delle due sequenze. Queste distanze vengono poi usate nel KNN per assegnare la sequenza ad una classe specifica.
L'implementazione del classificatore è stata realizzata utilizzando un package Python chiamato **Sequentia** (Onuonga (2019)). Questo pacchetto offre una gamma di algoritmi di classificazione e regressione per sequenze, tra cui metodi basati sul DTW (Dynamic Time Warping) e sui modelli nascosti di Markov. Il package fa uso dell'interfaccia di Scikit-Learn, riproponendo molti dei suoi moduli per facilitare la creazione di modelli. Tuttavia, sono state apportate modifiche all'interfaccia standard di Scikit-Learn per facilitare l'utilizzo di sequenze a lunghezza variabile. Come input il classificatore richiede tre tipologie di array:

- **Valori**: l'array bidimensionale contenente le sequenze temporali concatenate, dove ogni riga rappresenta un frame e ogni colonna rappresenta la specifica feature estratta da quel frame. Quindi, nel nostro caso, avremo un array bidimensionale in cui il numero di righe corrisponde alla somma dei frame estratti dai video e il numero di colonne pari a 15, che rappresenta il numero di feature estratte da ciascun frame.
- **Lunghezze**: un array dove ogni riga contiene un intero che rappresenta il numero di frame per ogni video fornito nell'array dei valori, così da poter determinare quanti frame utilizzare per ciascun video. Ciò ci permette di utilizzare video con lunghezze diverse, mantenendo allo stesso tempo un array bidimensionale di lunghezza fissa.
- **Etichette**: array contenente su ogni riga l'intero rappresentante le etichette assegnate ai video dell'array dei valori.

Per la creazione del classificatore, abbiamo utilizzato il modulo di scikit-learn chiamato sklearn.pipeline. La Pipeline ci permette di costruire una catena di funzioni per il processing dei dati e modelli di machine learning in un unico oggetto. La Pipeline da noi creata è costituita dai seguenti livelli ordinati:

1. **Denoise**: Rimozione del rumore da ogni sequenza applicando un filtro media per ogni feature;
2. **Scale**: Uso della funzione StandardScaler per la standardizzazione delle singole sequenze: sottraendo la media e dividendo per la deviazione standard di ciascuna feature. In questo modo, tutte le feature avranno la stessa scala e la stessa varianza unitaria, facilitando il loro confronto;
3. **PCA**: Riduzione della dimensionalità dei dati usando la PCA;
4. **KNN**: Livello finale che passa i risultati trasformati in un classificatore KNN .

5. Risultati

In questa sezione forniamo i risultati raggiunti dal nostro algoritmo di classificazione.

Una volta standardizzati i dati e applicata la PCA mantenendo tutti i componenti, abbiamo eseguito un analisi dei dati per scoprire il numero ottimale di componenti per il nostro modello attraverso lo scree plot in figura 4.

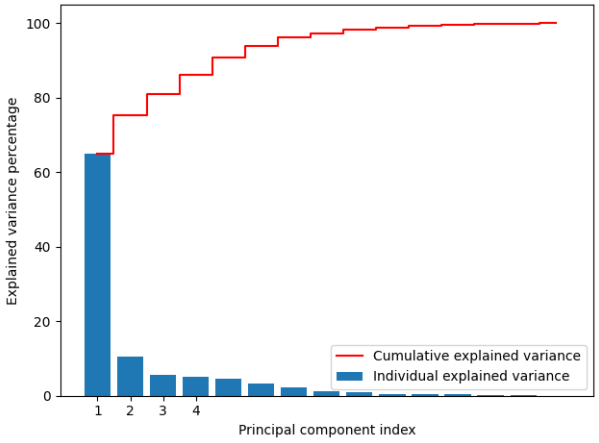

Figure 4: Scree plot varianza

Come si può evincere dallo scree plot, la prima componente principale riesce ad esprimere la maggior parte della varianza dei dati, per questo la scelta è ricaduta su un numero di componenti pari ad 1.

Dopo svariati test per la scelta ottimale dei parametri del classificatore K-nearest neighbors, si è giunti ad un valore di $\mathbf{k = 12}$ con una warping $\mathbf{window = 0.1}$. Questi valori hanno ottenuto un'accuratezza in fase di test pari al **67%**, con i valori visualizzabili nella tabella 4 e una matrice di confusione visualizzabile in figura 5.

Table 4: Tabella dei valori ottenuti in fase di test

	precision	recall	f1-score	support
Angry	0.33	1.00	0.50	1
Sad	1.00	0.43	0.60	7
Happy	0.75	1.00	0.86	3
Neutral	0.60	0.75	0.67	4
accuracy			0.67	15
macro avg	0.67	0.79	0.66	15
weighted avg	0.80	0.67	0.66	15

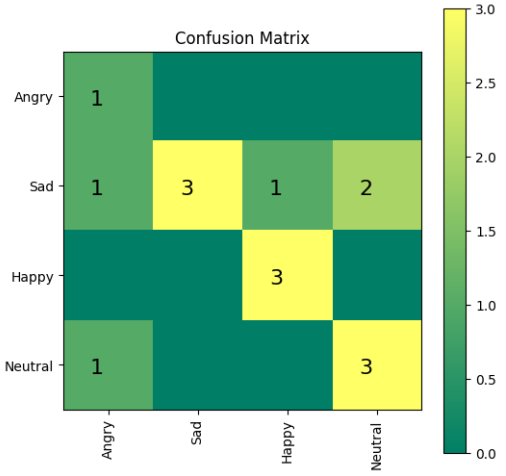

Figure 5: Matrice di Confusione

6. Discussione e Conclusioni

Il riconoscimento delle emozioni attraverso la camminata è un campo effettivamente inesplorato, principalmente a causa della scarsità di dataset disponibili e della mancanza di grandi quantità di dati, che possono influenzare le prestazioni di un modello e portare a risultati scarsi. Inoltre, la maggior parte dei video viene registrata in ambienti controllati, come citato nella sezione 2, che non riflettono le situazioni reali della vita di tutti i giorni, generando difficoltà nella creazione di un modello generale adatto al riconoscimento delle emozioni tramite l'andatura.

Un aspetto importante riguarda anche il processo di elaborazione dei dati. Gli algoritmi di pose estimation non sono dotati di un sistema di tracking, il che può portare a risultati inconsistenti in cui le coordinate dei punti di riferimento possono fluttuare, scambiarsi di posizione o sovrapporsi tra frame consecutivi. Inoltre, rispetto ai sistemi di tracking, vi è anche una differenza nel numero di keypoints estratti, ovvero quali punti di riferimento del corpo vengono considerati. Questo può influenzare la quantità e la qualità delle caratteristiche estratte. Ad esempio, nel lavoro di Jianwattanapaisarn, Sumi, Utsumi, Khamsemanan and Nattee (2022), viene utilizzato il sistema ben noto di motion capture chiamato *OptiTrack*, che estrae 37 markers e calcola 24 angoli per tutto il corpo.

Questo paper illustra un modello in grado di classificare l'emozione percepita da un soggetto che cammina in un video. L'approccio proposto consiste nell'applicazione di un algoritmo di pose estimation per determinare la posizione dei keypoints, dai quali vengono estratte manualmente 15 caratteristiche. Queste caratteristiche formano una sequenza temporale che viene utilizzata come input per il classificatore KNN, il quale utilizza la DTW (Dynamic Time Warping) per calcolare la distanza tra due oggetti anzichè consueta distanza euclidea. Dopo vari test per la scelta ottimale dei parametri, il modello raggiunge un'accuratezza pari al 67%.

References

Bhatia, Y., 2022. Bi-modal deep neural network for gait emotion recognition .

Castellano, G., Villalba, S.D., Camurri, A., 2007. Recognising human emotions from body movement and gesture dynamics, in: International conference on affective computing and intelligent interaction, Springer. pp. 71–82.

Chiu, M., Shu, J., Hui, P., 2018. Emotion recognition through gait on mobile devices , 800–805doi:10.1109/PERCOMW.2018.8480374.

Crenn, A., Khan, R.A., Meyer, A., Bouakaz, S., 2016. Body expression recognition from animated 3d skeleton, in: 2016 International Conference on 3D Imaging (IC3D), IEEE. pp. 1–7.

Jamil, N., Khir, N.H.M., Ismail, M., Razak, F.H.A., 2015. Gait-based emotion detection of children with autism spectrum disorders: a preliminary investigation. Procedia Computer Science 76, 342–348.

Jianwattanapaisarn, N., Sumi, K., Utsumi, A., Khamsemanan, N., Nattee, C., 2022. Emotional characteristic analysis of human gait while real-time movie viewing. Front. Artif. Intell. 5:989860 doi:10.3389/frai.2022.989860.

Kleinsmith, A., Bianchi-Berthouze, N., 2012. Affective body expression perception and recognition: A survey. IEEE Transactions on Affective Computing 4, 15–33.

Maji, D., Nagori, S., Mathew, M., Poddar, D., 2022. Yolo-pose: Enhancing yolo for multi person pose estimation using object keypoint similarity loss arXiv:2204.06806.

Mathivanan, B., Perumal, P., 2022. Recognition analysis for human identification analysis-a hybrid deep learning process. Wireless Pers Commun 126 , 555–579doi:https://doi.org/10.1007/s11277-022-09758-z.

Onuonga, E., 2019. sequentia. URL: https://github.com/eonu/sequentia#about.

Randhavane, T., Bhattacharya, U., Kapsaskis, K., Gray, K., Bera, A., Manocha, D., 2020. Identifying emotions from walking using affective and deep features arXiv:1906.11884.

Wallbott, H.G., 1998. Bodily expression of emotion. European journal of social psychology 28, 879–896.

Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M., 2022. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. arXiv preprint arXiv:2207.02696 .

Xu, S., Fang, J., Hu, X., Ngai, E., Wang, W., Guo, Y., Leung, V.C.M., 2022. Emotion recognition from gait analyses: Current research and future directions arXiv:2003.11461.