Emotion Gait

Foudal Soffian

ARTICLE INFO

Keywords:
emotion walk
landmark
emotion recognition
gait
machine learning

ABSTRACT

The dimensions of emotions significantly extend within the context of our existence, shaping our personal balance, social interactions, and the choices we make throughout our days. The complexity of emotions manifests through a broad range of internal states, such as the joy that brightens our experiences, the sadness that embraces us in moments of loss, the surge of anger that ignites our being, and many other nuances that enrich us.

In this investigation, we aim to explore the potential relationship between emotion and body movement during walking. Our goal is to pinpoint and identify traces of emotions within movement patterns, thus providing an opportunity for a deeper understanding of the emotional subtleties that permeate our daily lives.

The article is organized as follows: we will begin with an introductory overview that illustrates the context of emotions and their relevance in our existence. We will then delve into previous studies that have contributed to our current understanding, laying the groundwork for our research. Subsequently, we will examine in detail the implementation strategies we have adopted, including the learning models used for data analysis. We will present the results of our analyses and draw significant conclusions. Finally, we will offer future perspectives on the potential development of this line of research.

Through this research, we seek to contribute significantly to the understanding of human emotions and their impact on body movement during walking. We hope that our findings can open new horizons of knowledge and provide fresh insights in various fields, from psychology to artificial intelligence, thus enriching our understanding of emotions and their role in our lives.

1. Introduction

Emotions play a fundamental role in our lives, shaping our experiences, influencing our perspective of the world, and guiding our interactions with others. Being able to perceive the emotions of a person we are interacting with helps us understand their behavior and emotional state, enabling us to adopt an appropriate approach toward them.

Indeed, our response to a person often depends on the emotion they convey. If someone is sad, we seek to comfort them; if they are angry, we try to calm them, and so forth. Furthermore, the emotions of unfamiliar individuals can also influence our behavior. For instance, if we see a terrified or happy pedestrian on the street, our actions may be influenced by these conveyed emotions.

Given the significance of emotions, the automatic recognition of emotions represents a crucial challenge in various sectors, such as games, entertainment, and security. Typically, humans perceive emotions through verbal and body language. Recently, there has been an attempt to simulate this process through devices that, leveraging artificial intelligence, seek to understand emotions by analyzing verbal language and a person's behaviors.

The goal of this work is to determine if it is possible to perceive the emotion experienced by a person by analyzing their way of walking. This approach represents an additional method of emotion recognition without altering the initial research direction.

2. Related Work

In this chapter, we will examine related works that have contributed to the development of research on emotion identification based on gait. Two significant studies will be considered as the primary bibliographic references.

One of these studies, conducted by Tanmay Randhavane et al. [1], proposed a model based on deep features learned through LSTM to identify emotions perceived from people's gaits. Using an RGB video dataset, researchers extracted three-dimensional poses of individuals while walking and combined them with affective features calculated from gaits. The approach achieved an accuracy of 80.07

A second study, conducted by Venkatraman Narayanan et al. [2], focused on predicting emotions for socially aware robot navigation among pedestrians. Using gaits as input, the proposed model achieved an average accuracy of 82.47

Both works made use of the Mediapipe library [3] for extracting poses from video image frames.

These studies represent significant contributions in developing models for emotion identification based on gait. In our work, we will draw inspiration from these studies to further delve into our research.

3. Dataset

In this section, we will provide a description of the dataset used to conduct the project. The dataset consists of two parts: a collection of walking people videos and a CSV file.

Videos: A portion of the dataset comprises 60 videos. Each video displays a person walking frontally for a few seconds. During the pre-processing phase, some peculiarities in the videos emerged, posing various challenges:

- **Loop:** In each video, the same walking cycle is repeated multiple times, generating a series of loops. This characteristic required additional processing to extract a single loop from each video.
- **Censored Face:** Each video features a person with a censored face. This aspect caused some difficulties in pose extraction, as poses often appeared inverted. The lack of facial information made it challenging for the library to determine whether the person was facing front or back.
- **Reduced Size:** Another significant challenge of the dataset is its reduced size. Being composed of only 60 videos, the dataset is relatively small. This limited amount of data may impact the model's generalization and its ability to learn representative features accurately.

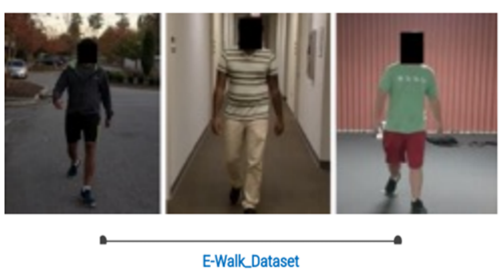


Figure 1: Some frames taken from dataset videos to provide a visual idea of these features.

CSV: The second part of the dataset consists of a CSV file structured as follows:

- The columns represent the walking people videos, and for each video, four questions are addressed:
  1. Is the person in the video happy?
  2. Is the person in the video angry?
  3. Is the person in the video sad?
  4. Is the person in the video neutral?

  Therefore, for each video, there are four columns, totaling 240 columns in the CSV file.
- The rows represent the participants who were asked the questions. When participants respond, they assign a rating on a scale from 1 to 5, where values have the following meanings:
  1. Strongly Disagree
  2. Somewhat Disagree
  3. Neither Agree nor Disagree
  4. Somewhat Agree
  5. Strongly Agree

It is important to emphasize that the information in the CSV file provides a subjective assessment of the emotions associated with walking video provided by survey participants.

4. Implementation Strategies

In this section, we will present the implementation strategy adopted to conduct the project.
To carry out the work, an approach involving the following phases was followed:

1. **Extraction of Statistics from the CSV File and Videos:** Statistical information was extracted from the CSV file, and video tracks were processed.
2. **Isolation of a Single Loop from Videos:** It was necessary to isolate a single walking cycle from each video, considering the presence of multiple loops.
3. **Extraction, Filtering, and Correction of Poses:** Poses were extracted from the videos, and filtering techniques were applied to improve data quality. Corrections were also made when necessary.
4. **Extraction of Features:** Relevant features were extracted from the poses to effectively represent the expressed emotion information.
5. **Application of Key Machine Learning Models:** The most relevant machine learning models were applied to analyze the extracted features and predict the emotion associated with walking.
6. **Analysis of Results:** The obtained results were analyzed and evaluated to assess the effectiveness of the machine learning model in emotion detection through walking.

It is important to emphasize that the implemented strategies were customized for the specific project and took into account the specific needs and requirements of the data processing and emotion modeling phases.

4.1. Extraction of Statistics from the CSV File and Video Labeling

In the initial phase of the project, some statistics based on participants' ratings in the CSV file were collected. In particular, averages and variances for each evaluated emotion were calculated. Figure 2 shows the proportions of emotions in the dataset, using the emotion with the highest average as the label.

Subsequently, Figure 3 illustrates the proportions of the training dataset, using the emotion with the highest average and establishing variance thresholds to handle neutral labels.

It is essential to note that using the average to generate labels led to an imbalance in the dataset in favor of the neutral emotion. About half of the videos are classified as neutral. However, considering the high variance associated

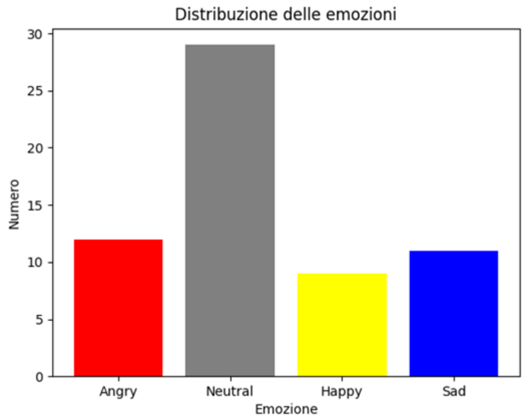


Figure 2: Emotion Averages

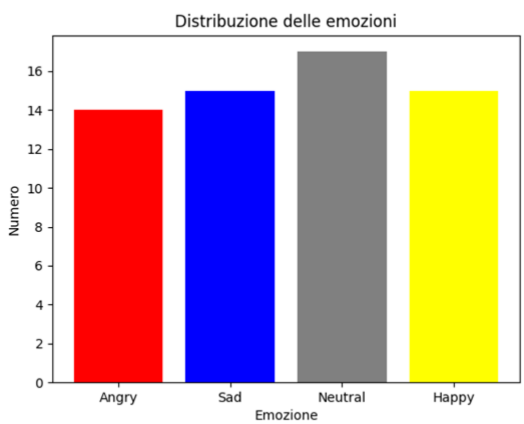


Figure 3: Emotion Variance Averages

with this emotion, a maximum variance threshold (1.25) was established for label generation.

This approach enables more accurate video labeling, taking into account variations in participant responses and ensuring a balanced representation of emotions in the dataset.

4.2. Isolation of a Single Loop from Videos

After analyzing the CSV file, we focused on video analysis. We encountered an initial challenge related to the presence of multiple loops, i.e., the repetition of the same walking cycle within each video. Therefore, our goal was to isolate a single walking cycle from each video.

The technique used is based on comparing video frames. The code analyzes frames one by one and looks for significant differences between them. To do this, it compares each frame with an initial reference frame.

If the sum of differences between the pixels of the current frame and those of the reference frame exceeds a predefined threshold, it is considered a significant change point in the video. These change points represent the boundaries between loops in the video.

In other words, the code identifies moments when the video transitions from one loop to another, detecting points where there are substantial differences between frames. These change points are used to divide the video into separate segments, each representing a single walking loop.

The ultimate goal is to obtain a video where only a single walking loop is retained, allowing for a more specific and focused analysis of a single movement cycle.

4.3. Pose Extraction and Filtering

In the third phase of the process, poses are extracted and filtered frame by frame. For pose extraction, the MediaPipe library was used, which detects landmarks representing the main parts of the body. Figure 4 shows the landmarks extracted from the library.

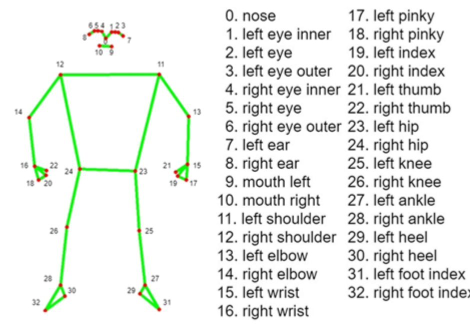


Figure 4: MediaPipe Landmarks

For our analysis, we chose not to consider some specific landmarks:

- Face Landmarks: Points 0 to 10 are not considered because faces in the provided videos are censored.
- Hand Landmarks: Points 17 to 22 are not considered due to the poor video quality, making it difficult for the library to correctly detect these points.
- Feet Landmarks: For the same reason mentioned above, foot landmarks from 29 to 32 are omitted.

After extracting poses, we noticed that some poses were not correctly identified for the following reasons:

- Incorrect Tracked Pose: Some tracked poses did not match the actual pose of the person in the video, as shown in the figure.
- Inverted Pose Orientation: Analyzing the coordinates of the extracted landmarks, we observed that in some frames, the right and left sides were inverted, indicating an error in the person's orientation. Figure 6 shows the coordinates of the landmarks frame by frame, highlighting the inversion in some frames.

To filter out incorrect poses, we checked if poses after the first frame experienced abrupt changes in shoulder and hip width measurements. If these variations exceeded a predefined threshold, the poses were not removed, as we would

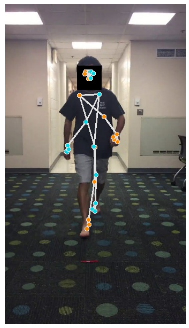


Figure 5: Incorrect Shoulder and Hip Pose

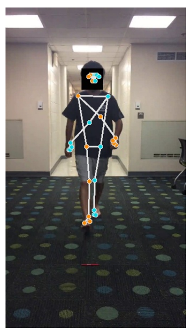


Figure 6: Inverted Poses

lose many frames in the process. However, adjustments were made to the values of the inverted pose to make it consistent with the last correctly recorded pose. The goal of this phase is to obtain correct and consistent poses that accurately represent body movements in the video.

4.4. Feature Extraction

Before proceeding with model identification, it is necessary to select the features that will be used for training. The following features have been chosen:

- Right arm length (right_arm_length)
- Left arm length (left_arm_length)
- Right leg length (right_leg_length)
- Left leg length (left_leg_length)
- Right elbow angle (right_elbow_angle)
- Left elbow angle (left_elbow_angle)
- Right knee angle (right_knee_angle)
- Left knee angle (left_knee_angle)
- Neck length (neck_length)
- Torso length (torso_length)
- Right leg to torso length ratio (right_leg_to_torso_ratio)
- Left leg to torso length ratio (left_leg_to_torso_ratio)
- Center of mass coordinates (centro_di_massa_x, centro_di_massa_y)

In addition to the above features, areas formed by figures with three body landmark points as vertices were also calculated. This approach captures additional information about the geometry and spatial distribution of body landmarks during walking. The areas can provide insights into the shape and pattern of body movements.

These features were selected to capture significant information about posture and body distribution during walking. Arm and leg lengths provide information about limb proportions, while elbow and knee angles indicate flexion of upper and lower limbs. Neck and torso length contribute to understanding body alignment, while ratios between leg and torso lengths highlight weight distribution during walking. Finally, center of mass coordinates provide information about body balance during movement. These features will be used as input for model training to detect and predict emotions associated with walking.

4.5. Application of Key Machine Learning Models

During the application of machine learning, several experiments were conducted to evaluate the performance of key models. This evaluation was done using 10-fold cross-validation. The models evaluated include Support Vector Machine, Decision Tree, Random Forest, and Logistic Regression.

To assess model performance, two different pipelines were used:

1. Evaluation of a pipeline on raw data, without any dimensionality reduction.

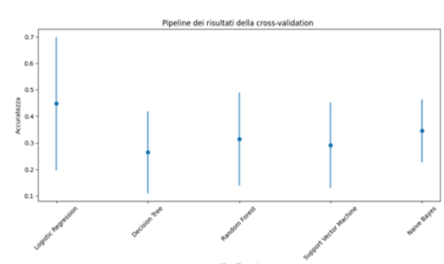


Figure 7: Pipeline without PCA

2. Evaluation of a pipeline where Principal Component Analysis (PCA) was applied to the data.

By using PCA, it was possible to reduce the dataset's dimensionality to 61 features.

These experiments allow for the comparison of model performance using different configurations and data pre-processing strategies. The goal is to identify which model and approach provide the best results in predicting emotions based on the features extracted from walking data.

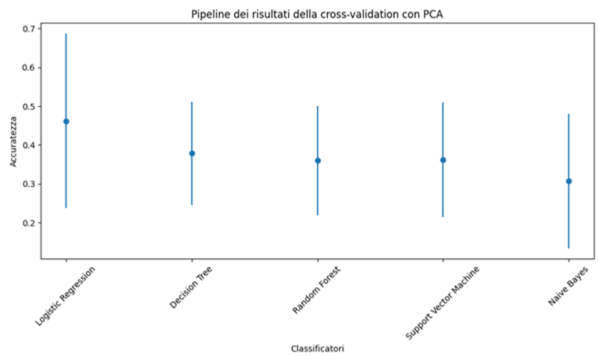


Figure 8: Pipeline with PCA

5. Results Analysis

In this chapter, we will analyze the results obtained in two configuration scenarios: original preprocessed data and preprocessed data resized using Principal Component Analysis (PCA).

5.1. Original Preprocessed Data

In the experiment using the original preprocessed data, we evaluated various configurations of key machine learning models using 10-fold cross-validation. The evaluated models include Decision Tree, Random Forest, and Logistic Regression.

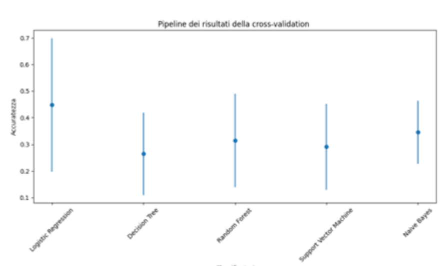
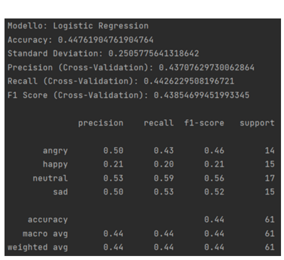


Figure 9: Pipeline results with cross-validation

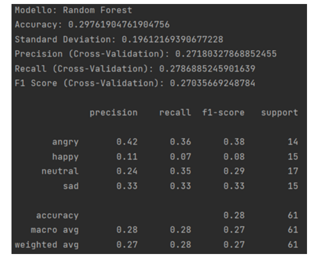
The results obtained indicate that the Logistic Regression model achieved the best performance in terms of accuracy, with a value of 45



```
Modello: Logistic Regression
Accuracy: 0.44761904761904764
Standard Deviation: 0.2505775641318642
Precision (Cross-Validation): 0.43707629730062864
Recall (Cross-Validation): 0.4426229508196721
F1 Score (Cross-Validation): 0.43854699451993345

              precision    recall  f1-score   support

       angry       0.50      0.43      0.46        14
       happy       0.21      0.20      0.21        15
     neutral       0.53      0.59      0.56        17
         sad       0.50      0.53      0.52        15

    accuracy                           0.44        61
   macro avg       0.44      0.44      0.44        61
weighted avg       0.44      0.44      0.44        61
```

Figure 10: Cross-validation results Logistic Regression

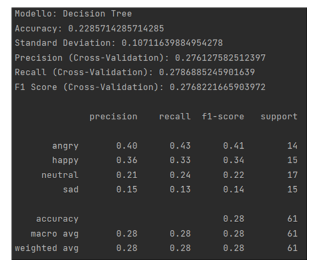
Subsequently, the Random Forest model achieved an accuracy of 30



```
Modello: Random Forest
Accuracy: 0.29761904761904756
Standard Deviation: 0.19612169390677228
Precision (Cross-Validation): 0.27180327868852455
Recall (Cross-Validation): 0.2786885245901639
F1 Score (Cross-Validation): 0.27035669248784

              precision    recall  f1-score   support

       angry       0.42      0.36      0.38        14
       happy       0.11      0.07      0.08        15
     neutral       0.24      0.35      0.29        17
         sad       0.33      0.33      0.33        15

    accuracy                           0.28        61
   macro avg       0.28      0.28      0.27        61
weighted avg       0.27      0.28      0.27        61
```

Figure 11: Cross-validation results Random Forest

While the Decision Tree model reached an accuracy of 23



```
Modello: Decision Tree
Accuracy: 0.2285714285714285
Standard Deviation: 0.10711639884954278
Precision (Cross-Validation): 0.276127582512397
Recall (Cross-Validation): 0.2786885245901639
F1 Score (Cross-Validation): 0.2768221665903972

              precision    recall  f1-score   support

       angry       0.40      0.43      0.41        14
       happy       0.36      0.33      0.34        15
     neutral       0.21      0.24      0.22        17
         sad       0.15      0.13      0.14        15

    accuracy                           0.28        61
   macro avg       0.28      0.28      0.28        61
weighted avg       0.28      0.28      0.28        61
```

Figure 12: Cross-validation results Decision Tree

5.2. Resized Preprocessed Data with PCA

In the experiment using the preprocessed data resized through PCA, we applied the same evaluation on machine learning models using 10-fold cross-validation.

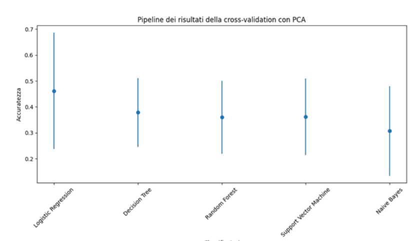
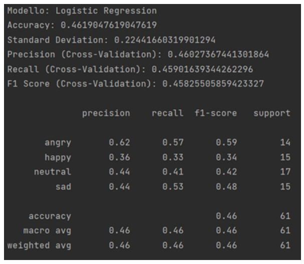


Figure 13: Pipeline results cross-validation and PCA

The results indicate that the Logistic Regression model showed the best performance again, reaching an accuracy of 46

Figure 14: Trend Logistic Regression

Subsequently, the Random Forest model achieved an accuracy of 36



Modello: Logistic Regression
Accuracy: 0.4619047619047619
Standard Deviation: 0.22441660319901294
Precision (Cross-Validation): 0.46027367441301864
Recall (Cross-Validation): 0.45901639344262296
F1 Score (Cross-Validation): 0.45825505859423327

	precision	recall	f1-score	support
angry	0.62	0.57	0.59	14
happy	0.36	0.33	0.34	15
neutral	0.44	0.41	0.42	17
sad	0.44	0.53	0.48	15
accuracy			0.46	61
macro avg	0.46	0.46	0.46	61
weighted avg	0.46	0.46	0.46	61

Figure 15: Results Logistic Regression

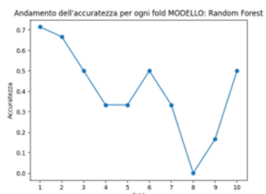
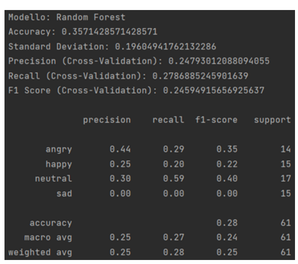


Figure 16: Trend Random Forest



Modello: Random Forest
Accuracy: 0.3571428571428571
Standard Deviation: 0.19604941762132286
Precision (Cross-Validation): 0.24793012088094055
Recall (Cross-Validation): 0.2786885245901639
F1 Score (Cross-Validation): 0.24594915656925637

	precision	recall	f1-score	support
angry	0.44	0.29	0.35	14
happy	0.25	0.20	0.22	15
neutral	0.30	0.59	0.40	17
sad	0.00	0.00	0.00	15
accuracy			0.28	61
macro avg	0.25	0.27	0.24	61
weighted avg	0.25	0.28	0.25	61

Figure 17: Results Random Forest

While the Decision Tree model reached an accuracy of 36

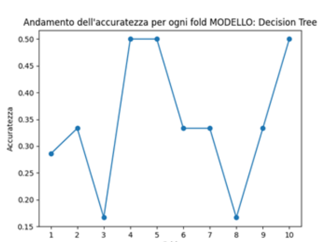
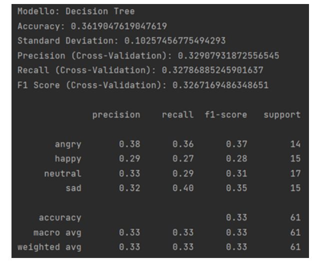


Figure 18: Trend Decision Tree

In both configurations, the Logistic Regression model has proven to be the most effective in predicting emotions based on the features extracted from walking data, achieving the highest accuracy. Figures 7 and 8 show the scores obtained by the models during cross-validation.

These results highlight the importance of data preprocessing and the effectiveness of the PCA technique in reducing the dataset's dimensionality. However, it is essential to consider that the achieved accuracy is still limited, and further model optimizations and refinements may be necessary to improve performance.



Modello: Decision Tree
Accuracy: 0.3619047619047619
Standard Deviation: 0.10257456775494293
Precision (Cross-Validation): 0.32907931872556545
Recall (Cross-Validation): 0.32786885245901637
F1 Score (Cross-Validation): 0.3267169486348651

	precision	recall	f1-score	support
angry	0.38	0.36	0.37	14
happy	0.29	0.27	0.28	15
neutral	0.33	0.29	0.31	17
sad	0.32	0.40	0.35	15
accuracy			0.33	61
macro avg	0.33	0.33	0.33	61
weighted avg	0.33	0.33	0.33	61

Figure 19: Results Decision Tree

6. Prediction Analysis on Two Test Sets

In this chapter, we will conduct a prediction analysis using two distinct test sets. The first test set consists of a reduced dataset of 15 videos that was provided to us earlier. The second test set is an extended dataset, including a larger number of videos for a more comprehensive evaluation of our prediction models' performance.

Our goal is to compare the performance of machine learning models on both test sets to assess the reliability and adaptability of the models to different data sizes.

We will start by running our pre-trained models on both test sets and record the results obtained for each model on both datasets. We will evaluate standard evaluation metrics such as accuracy, F1 score, precision, and recall to assess the models' performance on both test sets.

6.1. Reduced Test Dataset

In this chapter, we ran our pre-trained machine learning models to make predictions on a reduced dataset of 15 videos. Below, we present the results obtained for each model.

Our goal was to evaluate the performance of the following machine learning models:

- Logistic Regression
- Decision Tree
- Random Forest

We loaded the pre-trained models corresponding to each algorithm and used them to make predictions on the reduced dataset.

The results obtained are as follows: The Logistic Regression model achieved an accuracy of 40%, F1 score of 42%, precision of 44%, and recall of 40% (Figure 20).

Both the Random Forest and Decision Tree models achieved an accuracy of 53%, F1 score of 50%, precision of 48%, and recall of 53% (Figures 21 and 22).

6.2. Extended Test Dataset

Next, we report the results obtained with the three previously discussed models but on an extended test dataset.

For Logistic Regression, the results are the best, with an accuracy of 50%, F1 score of 52%, precision of 58%, and recall of 50% (Figure 23).

```
Modello in uso: Logistic Regression
Accuracy: 0.4
f1 score: 0.4205128205128205
precision: 0.4444444444444444
recall: 0.4
Matrice di confusione:
[[0 1 1 1]
 [0 4 2 1]
 [0 1 0 0]
 [0 0 2 2]]
classification report               precision    recall  f1-score   support

       angry       0.00      0.00      0.00         3
       happy       0.67      0.57      0.62         7
     neutral       0.00      0.00      0.00         1
         sad       0.50      0.50      0.50         4

    accuracy                           0.40        15
   macro avg       0.29      0.27      0.28        15
weighted avg       0.44      0.40      0.42        15
```

Figure 20: Results with Logistic Regression

```
Modello in uso: Random Forest
Accuracy: 0.5333333333333333
f1 score: 0.5037037037037037
precision: 0.48000000000000004
recall: 0.5333333333333333
Matrice di confusione:
[[0 2 0 1]
 [0 4 3 0]
 [0 1 0 0]
 [0 0 0 4]]
classification report               precision    recall  f1-score   support

       angry       0.00      0.00      0.00         3
       happy       0.57      0.57      0.57         7
     neutral       0.00      0.00      0.00         1
         sad       0.80      1.00      0.89         4

    accuracy                           0.53        15
   macro avg       0.34      0.39      0.37        15
weighted avg       0.48      0.53      0.50        15
```

Figure 21: Results with Random Forest

```
Modello in uso: Decision Tree
Accuracy: 0.5333333333333333
f1 score: 0.5023809523809525
precision: 0.4888888888888888
recall: 0.5333333333333333
Matrice di confusione:
[[0 2 0 1]
 [0 6 1 0]
 [0 1 0 0]
 [0 0 2 2]]
classification report               precision    recall  f1-score   support

       angry       0.00      0.00      0.00         3
       happy       0.67      0.86      0.75         7
     neutral       0.00      0.00      0.00         1
         sad       0.67      0.50      0.57         4

    accuracy                           0.53        15
   macro avg       0.33      0.34      0.33        15
weighted avg       0.49      0.53      0.50        15
```

Figure 22: Results with Decision Tree

```
Accuracy: 0.5
f1 score: 0.5166666666666666
precision: 0.5833333333333333
recall: 0.5
classification report               precision    recall  f1-score   support

                           angry       0.50      0.50      0.50         2
                           happy       0.33      0.50      0.40         2
                         neutral       0.50      0.50      0.50         2
                             sad       1.00      0.50      0.67         2

                        accuracy                           0.50         8
                       macro avg       0.58      0.50      0.52         8
                    weighted avg       0.58      0.50      0.52         8
```

Figure 23: Results with Logistic Regression

The Random Forest model achieved an accuracy of 50%, F1 score of 42%, precision of 37%, and recall of 50% (Figure 24).

The Decision Tree model yielded poor results with an accuracy of 25%, F1 score of 18%, precision of 15%, and recall of 25% (Figure 25).

```
Accuracy: 0.5
f1 score: 0.42500000000000004
precision: 0.375
recall: 0.5
classification report               precision    recall  f1-score   support

                           angry       0.00      0.00      0.00         2
                           happy       0.33      0.50      0.40         2
                         neutral       0.50      0.50      0.50         2
                             sad       0.67      1.00      0.80         2

                        accuracy                           0.50         8
                       macro avg       0.38      0.50      0.43         8
                    weighted avg       0.38      0.50      0.43         8
```

Figure 24: Results with Random Forest

```
Accuracy: 0.25
f1 score: 0.18333333333333335
precision: 0.14583333333333331
recall: 0.25
classification report               precision    recall  f1-score   support

                           angry       0.00      0.00      0.00         2
                           happy       0.25      0.50      0.33         2
                         neutral       0.00      0.00      0.00         2
                             sad       0.33      0.50      0.40         2

                        accuracy                           0.25         8
                       macro avg       0.15      0.25      0.18         8
                    weighted avg       0.15      0.25      0.18         8
```

Figure 25: Results with Decision Tree

7. Conclusions

The primary objective of this study was to investigate the possibility of extracting an individual's emotion solely by analyzing their way of walking. We presented the implementation strategies adopted, along with the results obtained in terms of accuracy, highlighting the accuracy trend graphs during the model training phase. The results were carefully analyzed, including a reflection on any obstacles encountered during the project's development.

In conclusion, the conducted experiments have shown that it is possible to detect a person's emotion by analyzing their way of walking. The developed neural network, validated on the reduced test set, achieved an accuracy of 54%. In particular, the models based on the Decision Tree and Random Forest algorithms, for validation on the extended test set, achieved an accuracy of 50%.

However, we believe that further improvement is possible by enhancing the quality of pre-processing and increasing the availability of data.

For future developments of this work, the following options could be considered:

- Acquiring a larger volume of data or applying deep learning techniques.
- Using a library for pose extraction where the presence of the face is not crucial to obtain more accurate poses.
- Using videos where the subject's face is not censored.

These potential research directions would allow further refinement of our approach and improvement of the overall performance of the proposed system.

References

[1] Tanmay Randhavane et al. "Identifying emotions from walking using affective and deep features". In: arXiv preprint arXiv:1906.11884v4 [cs.CV] 9 Jan 2020.

[2] Venkatraman Narayanan et al. "ProxEmo: Gait-based Emotion Learning and Multi-view Proxemic Fusion for Socially-Aware Robot Navigation". In: arXiv:2003.01062v2 [cs.RO] 28 Jul 2020.

[3] Mediapipe. URL: https://google.github.io/mediapipe/.