Comparison of Mediapipe and Yolo Pose detection to classify emotions

Bob Cruijsberg (852607441)
OU

Introduction

Emotion recognition is increasingly significant in fields like affective computing (Picard (2000)), impacting various domains such as education, healthcare, robotics, safety, and entertainment (Hernandez et al. (2021); Vinola and Vimaladevi (2015)). While traditional approaches often focused on speech, text, and facial expressions, recent studies are exploring the potential of detecting emotions through physiological signals and body postures (Garcia-Garcia, Penichet, and Lozano (2017); Saxena, Khanna, and Gupta (2020); Vinola and Vimaladevi (2015)). This work focuses on the latter, detecting emotions in body postures. The general strategy for emotion recognition across these modalities aligns with the framework outlined by Vinola and Vimaladevi (2015) as the emotion recognition framework (Figure 1)

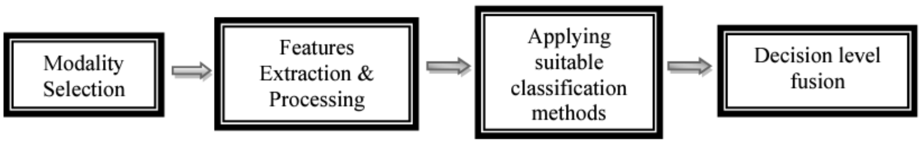


Figure 1: Emotion recognition framework

Figure 1. Emotion recognition framework (Vinola and Vimaladevi (2015))

With the fast-paced advancements in deep learning, significant progress in recognizing emotions from visual information has been made in the area of Feature Extraction & Processing. While the majority of research still concentrates on feature extraction from facial expressions, new deep learning methods are expanding to include body postures. These methods use silhouettes of body postures (Lee, Bae, Lee, and Kim (2017)), and focus on specific body parts or kinematic models (Blythe, Garrido, and Longo (2023); Noroozi et al. (2018); Saha, Datta, Konar, and Janarthanan (2014)) for feature extraction. The focus of this research will be on the kinematic models to extract features.

Goal

This research assesses the effectiveness of emotion classification using kinematic models, comparing feature sets retrieved from Mediapipe Pose landmarks and Yolov8 Pose keypoints. The goal is to identify which method provides better features for emotion recognition, by measuring and comparing the accuracy of the trained models. Such insights are expected to contribute valuable understanding to the field of emotion detection technology.

Research Question

How does the performance of classifying emotions vary when trained with different feature sets: one extracted from the Mediapipe Pose model and the other from the Yolov8 Pose model?

Null hypotheses

There is no significant difference in the performance of classifying emotion when trained with feature sets extracted from the Mediapipe Pose model compared to the Yolov8 Pose model.

To classify emotions, this study will utilize Ekman's six basic emotions and Neutral (Ekman and Friesen (1971)). The performance evaluation will be based on accuracy measurements to determine if there is a statistically significant difference.

The subsequent sections of this research is outlined as follows: The Data Analysis section covers the dataset's origin, limitations, deducting pose features from Mediapipe Pose/ Yolov8 Pose and an analysis. The Methodology section details the classifier optimization, and statistical testing. The Evaluation and Results sections analyse the outcomes and this is finally discussed in the Discussion section. From now on the Mediapipe Pose dataset will be refered to as the Mediapipe set and similar for the Yolov8 Pose Dataset as the Yolov8 set.

Data analysis

The initial data

This study uses 1124 images of 82 students (42 female) from the Ruhr-University Bochum, each displaying various emotions captured using a standard method (Thoma, Bauser, and Suchan (2013)). The emotions in the images were further validated with 19 students who identified the emotions from two options - the intended emotion and a random one. They also rated the naturalness of the emotions on a 1 to 5 scale. These validation results were compiled into a CSV file and linked to the respective images. Data cleaning was necessary as some images mentioned in the CSV file were not found in the image folder and vice versa. Additionally, there were some spelling errors in the image names in the CSV file. The dataset is furthermore balanced since it has an equal amount of images for each emotion.

Retrieving pose features using Yolov8 Pose and Mediapipe Pose

In the next step the Yolov8 Pose and Mediapipe Pose algorithms were used to detect respectively the Keypoints and Landmarks in the images, resulting in two different sets. For both the The results of this detection phase was filtered to only include images in which for Mediapipe the parameter Visibility was set to larger then 0.8, resulting in 1069 images,

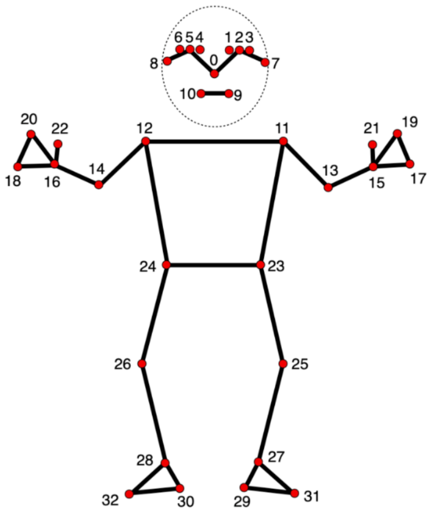


Figure 2. Mediapipe Landmarks

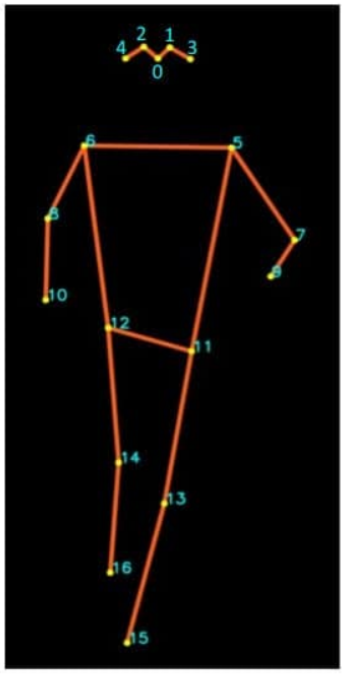


Figure 3. Yolov8 Keypoints

and for Yolov8 the property conf was also set to larger then 0.8, resulting in 1088 images. Both these parameters evaluate the accuracy of the detected Keypoints or Landmarks.

To convert the retrieved coordinates into useful features, angles were calculated according to the method described by Siam et al. (2022). A similar approach was used by Ferres, Schloesser, and Gloor (2022), they also retrieved the angles, but of dog postures. Figure B visually illustrates the method of angle calculation, and the formula used is as follows:
$\theta = \alpha - \beta = \tan^{-1}(\frac{y_3 - y_2}{x_3 - x_2}) - \tan^{-1}(\frac{y_1 - y_2}{x_1 - x_2})$
All angles were calculated (all 3 point combinations can be found in appendix B, for both the Yolov8 detection as for the Mediapipe detection and added as features to the datasets.

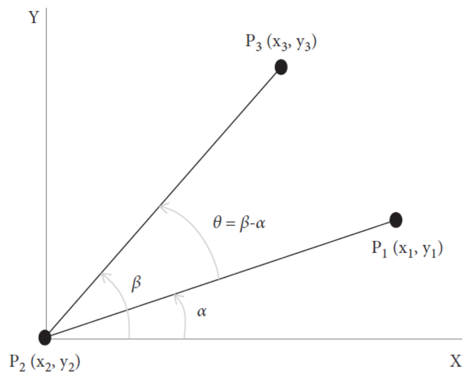


Figure 4. Calculating the angles (Siam et al. (2022))

Dataset analysis

In the previous paragraph, it was mentioned that detections with a confidence threshold lower than 0.8 were excluded. Next, considering the results of the validation session by Thoma et al. (2013), the distributions of two key variables were analyzed: the frequency of correct emotion recognition (Figure 5) and the perceived naturalness of the emotion (Figure 6). Based on these distributions, a decision was made to filter out data where Cor_cat_rate was less than 0.6 and Med_nat_rate was below 2.5. This filtering strategy ensured the retention of a sufficient amount of data for training purposes while also excluding data containing null values that failed to detect keypoints or landmarks.

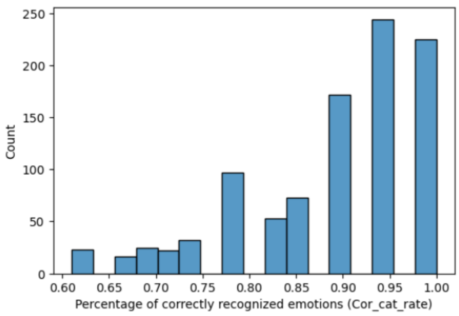


Figure 5. Histrogram of the Percentage of correctly recognized emotions

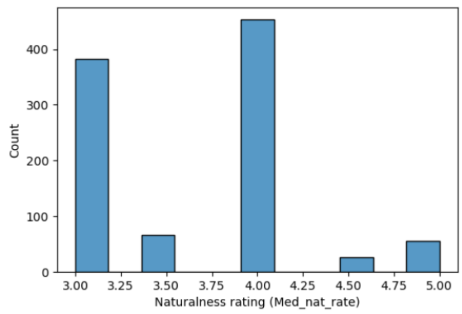


Figure 6. Histogram of the Naturalness ratings

In the next step, correlation matrices identified highly correlated angles, prompting the exclusion of redundant ones. This ensured a lighter, more efficient model with fewer features. The angles individually did not correlate much with the target variable Emotion. The following angles were removed:

For the Yolov8 set (see Figures 7, 8, 9)

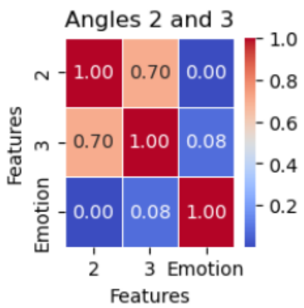


Figure 7. Correlation of Yolov8 Angle features 2 and 3

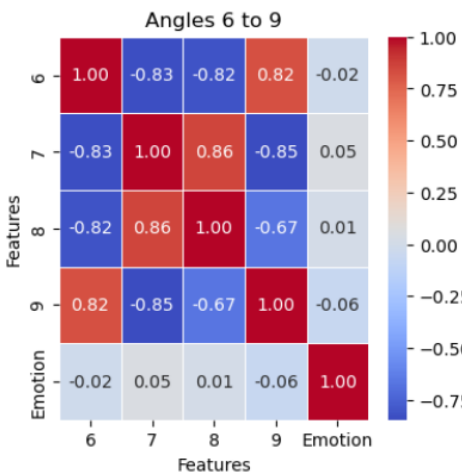


Figure 8. Correlation of Yolov8 Angle features 6 to 9

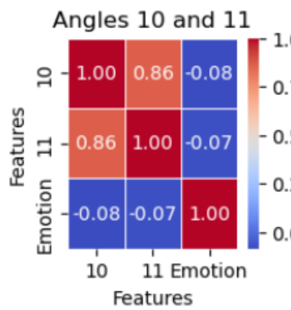


Figure 9. Correlation of Yolov8 Angle features 10 and 11

- Angles 2 and 3 both represent the torso. Angle 2 was removed for redundancy.
- Angles 6, 7, 8, and 9 all relate to the lower torso and its connection to the legs. Angles 6 and 7 were removed as angles 8 and 9 more effectively represent the position of the upper legs.
- Angles 10 and 11 determine the position of the legs. Both were retained as different leg positions could be significant for certain emotions.

The same was done for the Mediapipe set, the results can be found in appendix C.

Finally, the Image name and Facet variable were removed, the latter one determined the person's position: Averted or Frontal, which correlated with many of the Angle features.

Methodology and Implementation

Preparing the datasets

To reduce the number of features even more, PCA components were calculated. The aim of PCA components is to capture as much of the variance in the original data as possible, while at the same time reducing the number of features. Siam et al. (2022) also reduced the number of features by calculating PCA component, which improved their models. Disadvantage of using PCA components is that the results are less interpretable. To determine the best number of components the variance of each component (percentage) and the cumulative variance was plotted (see Fig D1 and in appendix D)

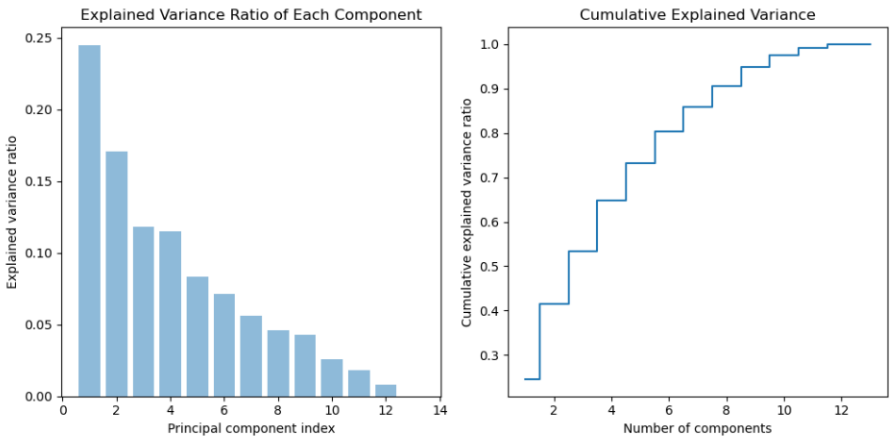


Figure 10. Variance and cummulative variance for the Yolov8 Pose dataset

For the MediaPipe set, 8 components were selected, and for the YOLOv8 set, 7 components were chosen. This selection accounted for at least 80% of the variance in both datasets. The initial graphs illustrate diminishing steps in variance reduction. Additionally, some models were calculated without this step, resulting in lower performance.

Furthermore, categorical data was dummified, floats were standardized, and the target value "Emotion" was encoded into numbers.

Selecting and training the models

SVM, KNN, Random Forest, and Logistic Regression were tested on a range of parameter combinations using GridSearchCV, which also employed cross-validation over 10 folds.

- SVM is a classifier that uses a hyperplane to separate the data points of two classes. In its basic form, it's a binary classifier, but it adopts an OvO (one versus one) approach for multiclass problems, testing all combinations of two classes. Important parameters include 'C', the regularization parameter, and the 'gamma' parameter (for an RBF kernel). Gamma values below 1 are chosen since a small gamma value defines a large similarity radius, meaning that points farther away from the decision boundary are considered in the calculation. This can lead to a more generalized model. The same principle applies to the 'C' parameter; smaller values are chosen, tolerating more errors, which further generalizes the model.
- K-Nearest Neighbors (KNN) classifies a new data point based on the categories of its nearest neighbors in the feature space. The algorithm measures closeness based on distance; in our case, the Euclidean distance was chosen as a measure (which is the default). Its most important parameter is 'k', the number of nearest neighbors. If it's too small, the model will likely overfit, and if it's too large, it may underfit.
- Random Forest is an ensemble learning method that builds multiple decision trees and merges their predictions. It uses bagging to train the different trees by bootstrapping various samples from the training data. Important parameters are 'n_estimators', determining the number of trees to be trained, and 'max_depth', determining the depth of a tree. Especially the latter should not be too large to avoid overfitting.
- Logistic Regression, like SVM, is by default a binary classifier, but in contrast to SVM, it uses an OvR approach (one versus the rest), creating binary problems for each class versus the other classes. Another option is to choose the Multinomial, but this is not supported by the liblinear solver, a solver which works well for multiclass classification. This is why for this study, the default lbfgs and liblinear were used. Furthermore, the penalty parameter was set to the default L2, as this is the only supported value for a multi-class classifier.

Finally a voting classifier was employed using all the previously mentioned models as an estimator, with the parameters set to the optimal values (based on the previous experiment). This is an ensemble method that uses voting to determine the final predictions.

Generating optimized models and testing the null hypotheses

The top-performing models will ultimately undergo training 500 times using varied training and testing samples, achieved by altering the Random State values. For each model, bagging will be implemented. This bootstrapping technique samples the training data multiple times with replacement, and aims to reduce variance. Following the training of the models on both datasets, the Mediapipe set and Yolov8 set, t-tests were used to evaluate the null hypothesis.

Evaluation and Results:

Overall mode evaluation

In the first step 5 different models were compared to each other (see Table 1). All models were tuned using a set of parameters as described in the method section. In the Mediapipe set, SVM, KNN and the Voting classifier performed overall best, although both SVM and the Voting classifier clearly struggled with overfitting. In the Yolov8 set it was mostly SVM that performed well, in this case also less overfitted as in the Mediapipe set.

Model	Mediapipe Pose		Yolov8 Pose	
	Training	Testing	Training	Testing
SVM	0.666	0.503	0.595	0.520
KNN	0.554	0.503	0.532	0.450
Random forest	0.668	0.462	0.684	0.455
Logistic Regression	0.424	0.442	0.430	0.425
Voting Classifier	0.617	0.508	0.599	0.475

Table 1
Model Comparison

In the next phase, we randomized the training/test set split 500 times and introduced a bagging classifier to enhance model generalizability. Due to the comparatively lower performance of Random Forest and Logistic Regression, our focus will shift away from these models. Similarly, although the Voting Classifier matched SVM's performance on the Mediapipe set, its complexity and lack of superiority led to its exclusion from further consideration. Thus, our attention will now be only on SVM and KNN models.

For each of the 500 models trained across various datasets, we conducted further analysis on those that were closest to the mean accuracy. As evident in Table 2, SVM models continue to show signs of overfitting, more so than KNN models. Upon examining the accuracies, it's observed that the best models correctly predict emotions approximately 49% of the time. While this accuracy is not exceptionally high, it significantly surpasses the likelihood of correctly guessing an emotion at random, which stands at about 14.3% (1 in 7 chance).

Model	Mediapipe Pose		Yolov8 Pose	
	Training	Testing	Training	Testing
SVM	0.699	0.496	0.610	0.491
KNN	0.566	0.489	0.545	0.455

Table 2
Model Comparison of model closes to the mean accuracy of 500 models

Performance of the different emotions

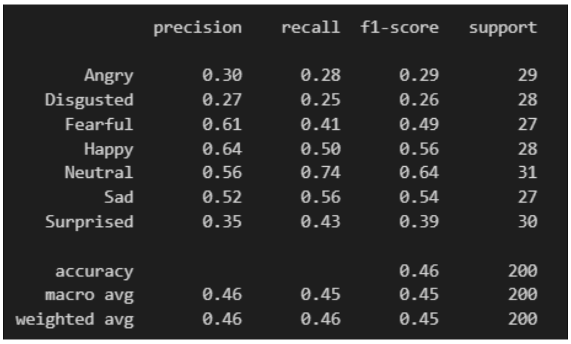
Looking at the f1 scores for the different target classes, we notices that some emotions are recognized much better than others (see for the Yolov8 set see figures 11 and 12 and for the Mediapipe set appendix E). Below the formules for Precision and Recall.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \tag{1}$$

As it becomes clear, Precision is mostly determined by the number of False Positives, in this case the KNN model for example in figure 13 predicts 14 times Angry, while these were actually 7 times surprised and 7 disgusted.

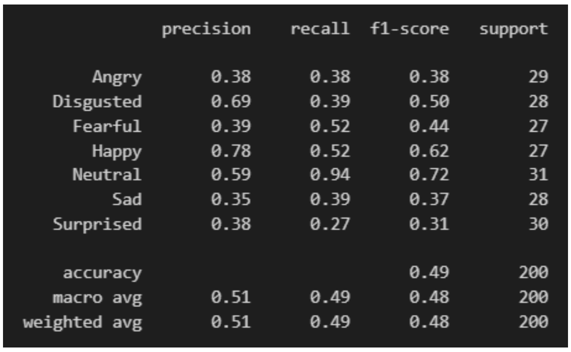
$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \tag{2}$$

The recall is mostly determined by the number of False Negatives, the number of times it doesn't find the correct emotion. For Disgusted the recall in 13 is low due to the fact that it often doesn't recognize Disgusted but thinks it is Angry or Surprised.



	precision	recall	f1-score	support
Angry	0.30	0.28	0.29	29
Disgusted	0.27	0.25	0.26	28
Fearful	0.61	0.41	0.49	27
Happy	0.64	0.50	0.56	28
Neutral	0.56	0.74	0.64	31
Sad	0.52	0.56	0.54	27
Surprised	0.35	0.43	0.39	30
accuracy			0.46	200
macro avg	0.46	0.45	0.45	200
weighted avg	0.46	0.46	0.45	200

Figure 11. Classification report KNN Yolov8 Pose



	precision	recall	f1-score	support
Angry	0.38	0.38	0.38	29
Disgusted	0.69	0.39	0.50	28
Fearful	0.39	0.52	0.44	27
Happy	0.78	0.52	0.62	27
Neutral	0.59	0.94	0.72	31
Sad	0.35	0.39	0.37	28
Surprised	0.38	0.27	0.31	30
accuracy			0.49	200
macro avg	0.51	0.49	0.48	200
weighted avg	0.51	0.49	0.48	200

Figure 12. Classification report SVM Yolov8 Pose

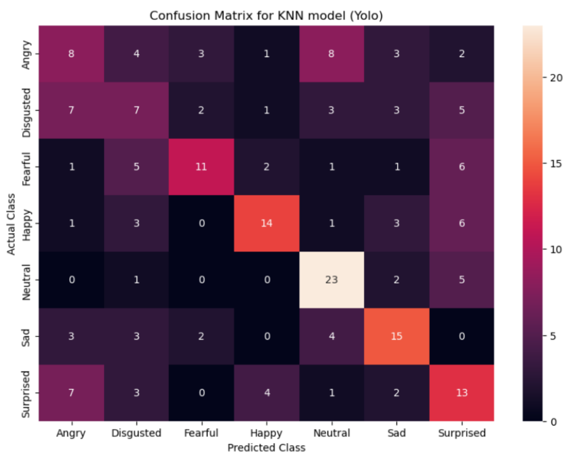


Figure 13. Confusion Matrix KNN Yolov8 Pose

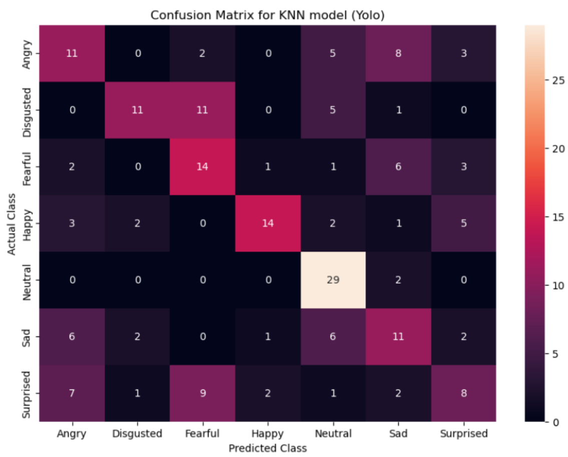


Figure 14. Confusion Matrix SVM Yolov8 Pose

Some interesting observations:

- The classification of the emotions 'Disgusted' and 'Angry' often leads to confusion across classes, with the notable exception of the Yolov8 SVM model. This model uniquely distinguishes these two emotions without error.
- The Yolov8 SVM model also stands out in several aspects: it frequently achieves a high precision score for 'Disgusted', implying it rarely misidentifies other emotions as 'Disgusted'. However, it demonstrates a notably low recall score for 'Surprised', indicating a tendency to overlook this emotion in classifications.
- In all models, 'Neutral' is consistently well-recognized, particularly in terms of recall, suggesting that almost all neutral images are correctly identified.
- Additionally, 'Happy' tends to be more accurately classified in the Mediapipe models, indicating their enhanced effectiveness in recognizing this specific emotion.

Null hypothesis testing

After analyzing the 500 models for each classifier and testing their significance relative to each other, it is observed in Figure 15 that three sets of models nearly completely overlap, suggesting similar prediction capabilities, with the exception of the KNN Yolov8 models. This observation might imply that there is no significant difference between most of the models, leading to the acceptance of the null hypotheses. However, upon examining the confidence intervals on the SVM plots from both datasets, very narrow intervals are noted, which appear not to overlap (refer to Figure 16).

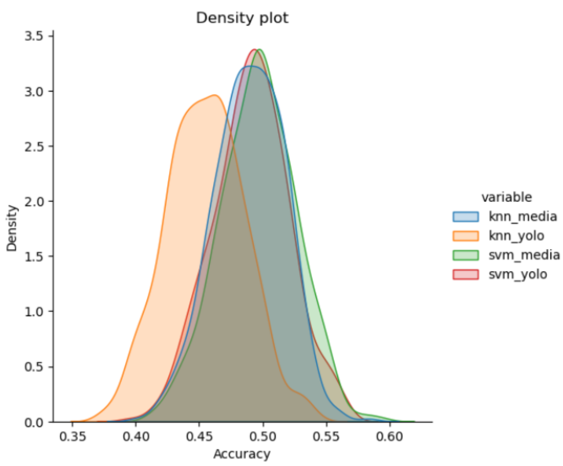


Figure 15. Density Plot for all 4 models

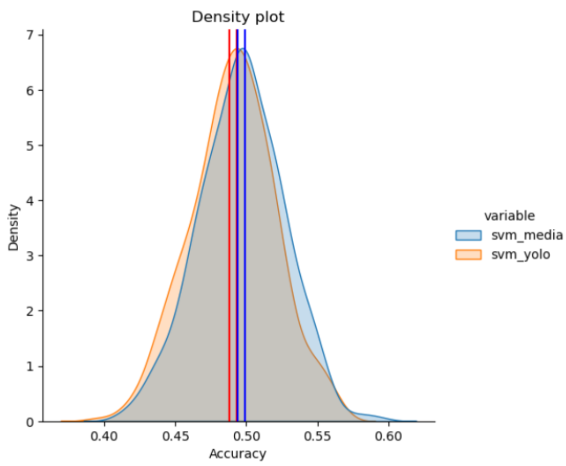


Figure 16. Density plot for the SVM models with Confidence interval lines (95%)

In Table 3, it shows that the Standard error is in all cases very small, resulting in very narrow Confidence intervals.

After these initial observations, t-tests were performed. The results, as depicted in Figure 4, indicate that in most instances, there is a significant difference between the

	Mean	Standard Error	Confidence Interval
KNN Mediapipe	0.489	0.001	0.486 - 0.493
KNN Yolov8	0.455	0.001	0.451 - 0.458
SVM Mediapipe	0.496	0.001	0.493 - 0.500
SVM Yolov8	0.491	0.001	0.487 - 0.494

Table 3
statistics of the 4 groups of data

models trained with the Mediapipe dataset and those trained with Yolo, except for the KNN models trained with Mediapipe data versus the SVM models trained with Yolov8 data. In this particular case, the null hypothesis is accepted, as the p-value exceeds 0.05, indicating no significant difference at a 95% confidence interval.

	T value	P value
KNN Mediapipe versus KNN Yolov8	18.736	0.000
SVM Mediapipe versus SVM Yolov8	3.050	0.002
KNN Mediapipe versus SVM Yolov8	-0.682	0.496

Table 4
Null hypothesis testing with ttest's

Conclusion and Discussion

This study aimed to explore the differences in performance between features derived from the Mediapipe Pose model and those from the Yolov8 Pose model in classifying emotions. Given that the Mediapipe Pose model detects twice as many points as the Yolov8 Pose model, it was anticipated that the latter might underperform. Although significant differences were observed in the performance of both KNN and SVM models, they were not that different. The KNN models using Yolov8 data and the SVM models trained with either Yolov8 or Mediapipe data exhibited average accuracies ranging from 0.489 to 0.496. However, due to the small standard error and thus narrow confidence intervals, only the KNN models trained with the Mediapipe data were found to be significantly comparable in performance to the SVM models trained with the Yolov8 data. For these models alone, the Null Hypothesis was accepted. In comparison to the studies by Lee et al. (2017) and Saha et al. (2014), the accuracies in this research were not as high.

Beyond similarities in accuracies, notable differences and similarities in emotion classification by the models were observed. Particularly, the SVM models trained with the Yolov8 data deviated in performance for more than one emotion compared to other models, which were more closely aligned. As expected, the more distinct emotion 'Happy' performed relatively well, whereas 'Disgusted' fared worse. Surprisingly, 'Neutral' performed well, especially in Recall, suggesting most Neutral images were correctly identified. Contrary to expectations, Neutral was assumed to be more challenging to distinguish as a separate emotion.

A limitation of this study was the computing power and time constraints, leading to the use of a limited number of models, minimal fine-tuning, and the exclusion of deep learning models. These factors could potentially enhance accuracies and align results more closely with those found in Saha et al. (2014). Additionally, the use of more data might improve performance and reveal larger differences between models trained with Yolov8 versus Mediapipe data.

Emotion detection also raises ethical concerns. Andalibi and Buss (2020) highlight that emotions, being intimate and personal, could be exploited for manipulation and are complex to define, as demonstrated in this study where the classifier correctly identified emotions only about 49% of the time. Hernandez et al. (2021) further underscore the risk of human rights violations when emotion detection systems are used in public spaces, schools, or companies, or to deny access to services like job opportunities, loans, or social services. To mitigate these risks, they proposed a set of guidelines, which could help apply the insights from this study to practical applications in an ethical and moral manner.

References

Andalibi, N., & Buss, J. (2020). The human in emotion recognition on social media: Attitudes, outcomes, risks. In *Proceedings of the 2020 chi conference on human factors in computing systems* (pp. 1–16).

Blythe, E., Garrido, L., & Longo, M. R. (2023). Emotion is perceived accurately from isolated body parts, especially hands. *Cognition*, *230*, 105260.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124.

Ferres, K., Schloesser, T., & Gloor, P. A. (2022). Predicting dog emotions based on posture analysis using deeplabcut. *Future Internet*, *14*(4), 97.

Garcia-Garcia, J. M., Penichet, V. M., & Lozano, M. D. (2017). Emotion detection: a technology review. In *Proceedings of the xviii international conference on human computer interaction* (pp. 1–8).

Hernandez, J., Lovejoy, J., McDuff, D., Suh, J., O'Brien, T., Sethumadhavan, A., … Czerwinski, M. (2021). Guidelines for assessing and minimizing risks of emotion recognition applications. In *2021 9th international conference on affective computing and intelligent interaction (acii)* (pp. 1–8).

Lee, S. K., Bae, M., Lee, W., & Kim, H. (2017). Cepp: Perceiving the emotional state of the user based on body posture. *Applied Sciences*, *7*(10), 978.

Noroozi, F., Corneanu, C. A., Kamińska, D., Sapiński, T., Escalera, S., & Anbarjafari, G. (2018). Survey on emotional body gesture recognition. *IEEE transactions on affective computing*, *12*(2), 505–523.

Picard, R. W. (2000). *Affective computing.* MIT press.

Saha, S., Datta, S., Konar, A., & Janarthanan, R. (2014). A study on emotion recognition from body gestures using kinect sensor. In *2014 international conference on communication and signal processing* (pp. 056–060).

Saxena, A., Khanna, A., & Gupta, D. (2020). Emotion recognition and detection methods: A comprehensive survey. *Journal of Artificial Intelligence and Systems*, *2*(1), 53–79.

Siam, A. I., Soliman, N. F., Algarni, A. D., El-Samie, A., Fathi, E., Sedik, A., et al. (2022). Deploying machine learning techniques for human emotion detection. *Computational intelligence and neuroscience*, *2022*.

Thoma, P., Bauser, D. S., & Suchan, B. (2013). Besst (bochum emotional stimulus set)—a pilot validation study of a stimulus set containing emotional bodies and faces from frontal and

averted views. *Psychiatry Research*, *209*(1), 98–109.
Vinola, C., & Vimaladevi, K. (2015). A survey on human emotion recognition approaches, databases and applications. *ELCVIA: electronic letters on computer vision and image analysis*, 00024–44.

Appendix A
Code and dataset

The initial dataset with images was downloaded from: `https://www.ruhr-uni-bochum.de/neuropsy/BESST.html`
All code can be found on Github: `https://github.com/bcruijsberg/ML-OU`

Appendix B
Combinations of Angles

Angle	Points	Angle	Points	Angle	Points	Angle	Points
0	(18,16,20)	5	(15,13,11)	10	(24,23,11)	15	(23,25,27)
1	(22,16,14)	6	(14,12,24)	11	(23,24,12)	16	(26,28,32)
2	(16,14,12)	7	(13,11,23)	12	(23,24,26)	17	(30,32,28)
3	(17,15,19)	8	(24,12,11)	13	(25,23,24)	18	(25,27,31)
4	(21,15,13)	9	(12,11,23)	14	(24,26,28)	19	(29,31,27)

Table B1
Angles Mediapipe

Angle	Points	Angle	Points	Angle	Points	Angle	Points
0	(10,9,6)	3	(6,5,11)	6	(6,12,11)	9	(13,11,12)
1	(8,6,12)	4	(11,5,7)	7	(5,11,12)	10	(12,14,16)
2	(12,6,5)	5	(5,7,9)	8	(14,12,11)	11	(11,13,15)

Table B2
Angles Yolov8

Appendix C
Correlation Mediapipe Pose

For the Mediapipe Pose dataset (see Figures C1, C2, C3)

- Angles 0 and 1 both represent angles of the pulse, similar to angles 3 and 4 for the other pulse. Therefore, angles 0 and 3 were removed.
- Angles 8 and 9 both relate to the torso. Angle 8 was removed.
- Angles 10, 11, 12, and 13 all define the lower torso and its connection to the legs. Angles 10 and 11 were removed as angles 12 and 13 more effectively represent the position of the upper legs.
- Angles 16 and 17, and 18 and 19, determine the foot's connection and width, respectively, for each leg. Angles 17 and 19, which determine the width of the foot, were removed.

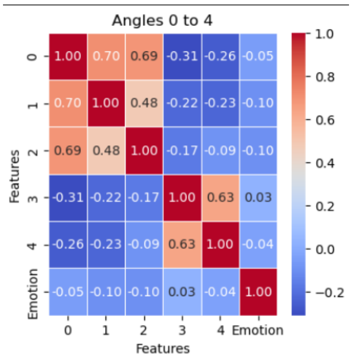


Figure C1. Correlation of Mediapipe Angle features 0 to 4

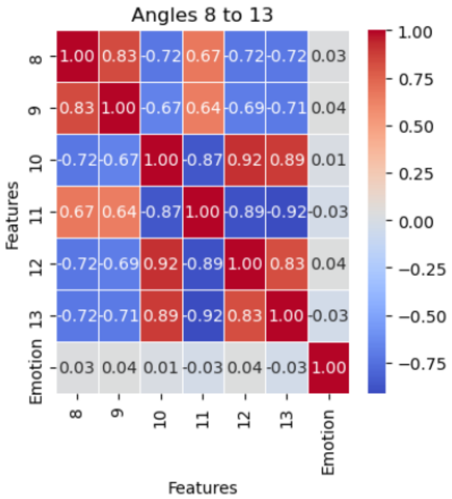


Figure C2. Correlation of Mediapipe Angle features 8 to 13

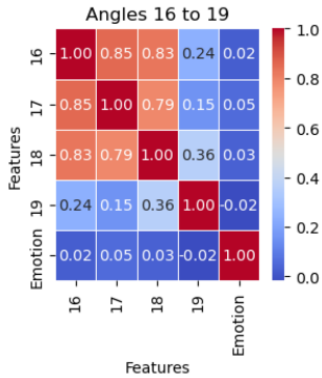


Figure C3. Correlation of Mediapipe Angle features 16 to 19

Appendix D
PCA variance Mediapipe Pose

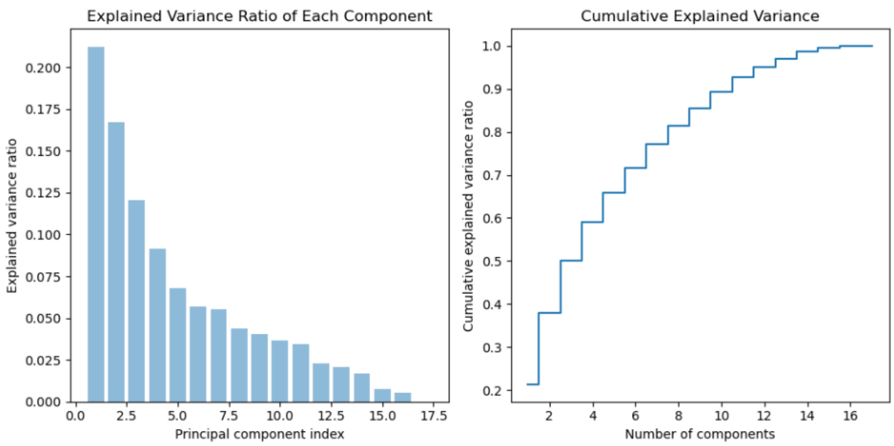


Figure D1. Variance and cummulative variance for the Mediapipe Pose dataset

Appendix E
results Mediapipe Pose

	precision	recall	f1-score	support
Angry	0.20	0.15	0.17	27
Disgusted	0.30	0.22	0.26	27
Fearful	0.52	0.50	0.51	26
Happy	0.68	0.72	0.70	29
Neutral	0.63	0.87	0.73	30
Sad	0.48	0.39	0.43	28
Surprised	0.41	0.50	0.45	30
accuracy			0.49	197
macro avg	0.46	0.48	0.46	197
weighted avg	0.46	0.49	0.47	197

Figure E1. Classification report KNN Mediapipe Pose

	precision	recall	f1-score	support
Angry	0.25	0.19	0.21	27
Disgusted	0.20	0.19	0.19	27
Fearful	0.39	0.50	0.44	26
Happy	0.83	0.69	0.75	29
Neutral	0.68	0.83	0.75	30
Sad	0.52	0.50	0.51	28
Surprised	0.52	0.53	0.52	30
accuracy			0.50	197
macro avg	0.48	0.49	0.48	197
weighted avg	0.49	0.50	0.49	197

Figure E2. Classification report SVM Mediapipe Pose

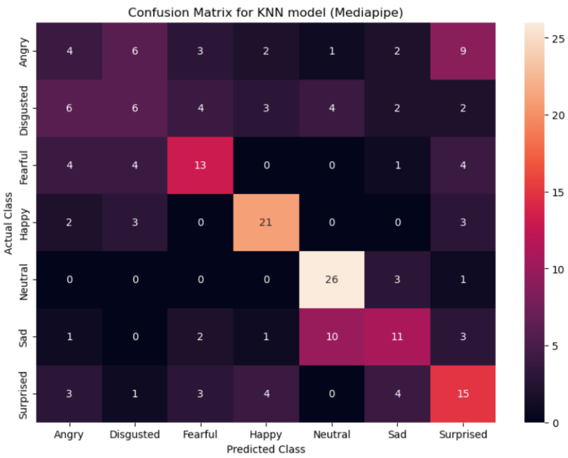


Figure E3. Confusion Matrix KNN Mediapipe Pose

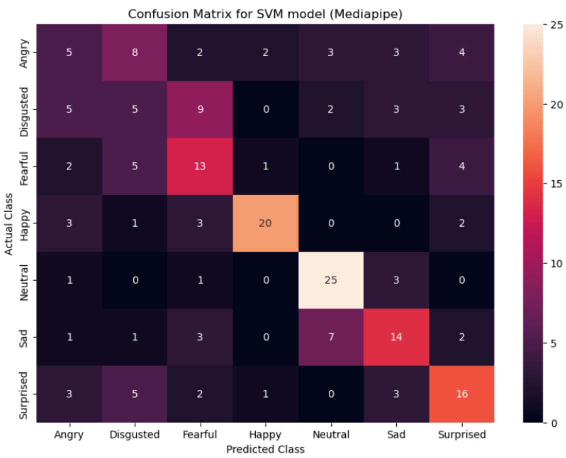


Figure E4. Confusion Matrix SVM Mediapipe Pose