Computer Vision Pipeline for Face Shape Classification using Facial Landmarks

Guilherme Duran Duran Gea, João Victor Guimarães Campelo and Laura Calleja García-Navas

***Abstract*— This work presents a computer vision pipeline for classifying face shapes into four categories using geometric features extracted from MediaPipe FaceMesh landmarks. Two strategies are evaluated: deterministic threshold rules and machine-learning models. Using a balanced dataset of 1,000 images per class, threshold rules show limited performance, while Random Forest and Gradient Boosting achieve substantially higher accuracy. The results indicate that geometric ratios provide useful but incomplete information, and that combining structured features with learning-based methods improves face-shape classification.**

***Keywords*— Face Detection, Landmark Detection, Face Shape Classification**

I. Introduction

The analysis of facial geometry supports applications in fashion, aesthetics, and computer vision. Identifying face shape assists in personalized recommendations such as hairstyles, beard designs, and makeup techniques. Computationally, face-shape classification provides a structured representation of facial proportions that can be quantitatively analyzed. Prior studies have shown that geometric and landmark-based representations can reliably support face analysis tasks, including detection and recognition [1] and even direct face-shape prediction for recommendation systems [2].

This work proposes a pipeline for classifying faces into four morphological categories: square, oval, round, and heart. The process includes face detection, landmark extraction, geometric feature computation, and classification using either deterministic rules or machine-learning models. Detection and landmark estimation rely on FaceMesh, a MediaPipe model that outputs 468 detailed landmarks. These serve as the basis for constructing interpretable geometric descriptors linked to facial proportions.

The remainder of this paper is structured as follows. Section III describes the pipeline in detail. Section IV presents and discusses the experimental results. Section V concludes the paper and highlights directions for future work.

II. Problem Statement and Assumptions

Face shape can be classified into four main categories: round, square, oval, and heart. Each shape can be identified using a set of geometric rules based on the relative proportions of the face:

(a) Square face (b) Round face

(c) Oval face (d) Heart face

Fig. 1. Examples of the four face shape categories used in this study: (a) square, (b) round, (c) oval, and (d) heart.

While rule-based geometric criteria provide interpretability and simplicity, they are also sensitive to variations in landmark precision and image conditions. Small errors in landmark detection or slight changes in face orientation can lead to misclassification.

To reduce these effects, we restrict our analysis to images that meet specific quality criteria:

- Proper and uniform illumination across the face;
- Clearly visible facial boundaries (forehead, jawline, and chin contours identifiable);
- The subject positioned at a distance of up to two meters from the camera;
- A neutral pose, with minimal head rotation or tilt.

III. Methodology

A. Face Detection and Landmark Extraction

The first stage of the proposed pipeline, shown in Fig. 2(a), employs the MediaPipe FaceMesh framework [3], which internally performs face detection before extracting facial landmarks. This integrated approach eliminates the need for a separate detection model, as FaceMesh initially localizes the facial region within the image and defines a bounding box to delimit the area of interest. A minimum confidence threshold of 0.75 was applied to ensure that only reliable

Guilherme Duran Duran Gea and João Victor Guimarães Campelo are undergraduate students of Data Science and Artificial Intelligence at IBMEC-RJ (Centro Universitário IBMEC, Rio de Janeiro, Brazil). Emails: guilherme.d.gea@gmail.com, jvguimaraescampelo@gmail.com.

Laura Calleja García-Navas is an undergraduate student of Artificial Intelligence at Universidad Rey Juan Carlos (Madrid, Spain) and is currently an exchange student at IBMEC-RJ. Email: lauracgn28@gmail.com.

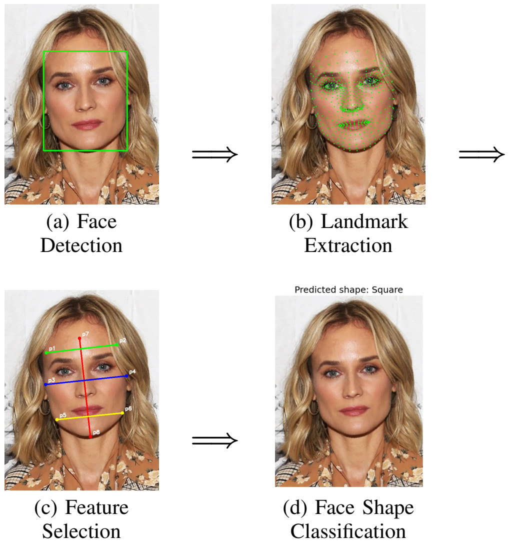

Fig. 2. Pipeline of the proposed facial shape classification system: (a) face detection, (b) facial landmark extraction, (c) feature selection, and (d) final face shape classification. The arrows indicate the processing flow between stages.

detections were considered, thereby improving the robustness of the pipeline under moderate variations in pose, illumination, and expression.

In the subsequent step, FaceMesh extracts 468 3D facial landmarks from the detected region, as illustrated in Fig. 2(b), providing a dense and detailed geometric representation of the face. These landmarks capture both coarse and fine-grained facial structures, serving as a solid foundation for higher-level tasks such as facial analysis and classification.

B. Feature Calculation

After the extraction of the eight selected landmarks, as shown in Fig. 2(c), the next step involves computing geometric relationships that quantitatively describe facial proportions. From the coordinates of the landmarks, we calculate four key distances representing distinct facial regions: facial height, forehead width, cheekbone width, and jawline width, as illustrated below:

$$d_f = \|p_1 - p_2\| \quad \text{(Forehead width)} \tag{1}$$

$$d_c = \|p_3 - p_4\| \quad \text{(Cheekbone width)} \tag{2}$$

$$d_j = \|p_5 - p_6\| \quad \text{(Jawline width)} \tag{3}$$

$$d_h = \|p_7 - p_8\| \quad \text{(Facial height)} \tag{4}$$

From these primary distances, a set of five ratios (R_i) was derived to represent relative facial proportions. These ratios serve both as interpretable indicators for manual threshold-based classification and as numerical features for the deep learning model:

$$R_1 = \frac{d_h}{d_c} \qquad \text{Height-to-cheekbone ratio} \tag{5}$$

$$R_2 = \frac{d_j}{d_f} \qquad \text{Jaw-to-forehead ratio} \tag{6}$$

$$R_3 = \frac{d_j}{d_c} \qquad \text{Jaw-to-cheekbone ratio} \tag{7}$$

$$R_4 = \frac{d_f}{d_c} \qquad \text{Forehead-to-cheekbone ratio} \tag{8}$$

$$R_5 = \frac{d_f}{d_j} \qquad \text{Forehead-to-jaw ratio} \tag{9}$$

These normalized ratios are invariant to face scale and image resolution, ensuring that the derived features capture the structural balance among facial regions rather than absolute dimensions. This property is essential for enabling robust comparisons across different subjects and image conditions.

C. Threshold-based Face Shape Classification

TABLE I. Geometric ratios and threshold ranges used for each face-shape category.

Face shape	$R_1 = \frac{d_h}{d_c}$	$R_2 = \frac{d_j}{d_f}$	$R_3 = \frac{d_j}{d_c}$	$R_4 = \frac{d_f}{d_c}$	$R_5 = \frac{d_f}{d_j}$
Heart	> 0.95	< 0.90	< 0.85	–	> 1.10
Oval	0.90–1.10	0.85–0.95	0.75–0.85	–	1.05–1.12
Round	≈ 1.00	–	–	≈ 0.90	1.10–1.20
Square	< 0.90	0.90–0.97	0.80–0.85	–	< 1.10

Initially, we attempted to classify the face shapes using fixed geometric thresholds based on the ratios derived from the facial distances defined in Equations 5–9. The criteria summarized in Table I were empirically defined from visual analysis of the dataset and the literature on facial morphology.

Although this threshold-based approach allowed for an interpretable and geometrically grounded classification, the resulting accuracy was limited due to natural variation in human facial proportions and landmark detection sensitivity. Therefore, this step served primarily as a preliminary analysis to guide the development of the machine learning-based model described in the following section.

D. Machine Learning-based Face Shape Classification

After computing the geometric ratios defined in Equations 5–9, we trained two supervised models—Random Forest and Gradient Boosting—using these features as inputs. The adoption of learning-based methods follows prior evidence that face-shape classification benefits from comparative evaluations across multiple algorithms [4]. These algorithms were chosen due to their strong performance in nonlinear tabular classification tasks, with Random Forest offering robust multi-class behavior through ensemble averaging [5] and Gradient Boosting being recognized for its high predictive accuracy and iterative error correction [6]. Both models were evaluated under the same conditions as the threshold-based approach, enabling a direct comparison between rule-driven and learning-based strategies.

E. Dataset

The dataset used in this study is based on the publicly available *Face Shape Dataset*, which originally consists of approximately 5,000 frontal images of female celebrities. For the purposes of this work, we selected a subset comprising four face shape categories relevant to our analysis, with 1,000 images per category. Each image is labeled according to its facial morphology.

IV. Results and Discussion

This section presents the comparative evaluation of the threshold-based classifier introduced in Section III-C and the machine-learning models described in Section III-D. All methods were assessed using accuracy and confusion matrices across the four face-shape categories.

A. Threshold-Based Classification Results

The threshold-based method achieved an overall accuracy of 20.25%, demonstrating that fixed geometric rules derived from ratios 5–9 were insufficient for reliable face-shape discrimination. Human facial proportions vary substantially, and handcrafted thresholds are highly sensitive to small deviations in landmark localization. As a result, the classifier frequently collapsed predictions into a dominant class.

Figure 3 shows the confusion matrix for this method. The strong concentration of predictions in a single category highlights the limited discriminative power of rule-based geometric heuristics when applied to fine-grained facial shape categories. Misclassifications remain frequent among visually similar categories, especially between Oval and Heart, reflecting inherent ambiguities in human facial morphology.

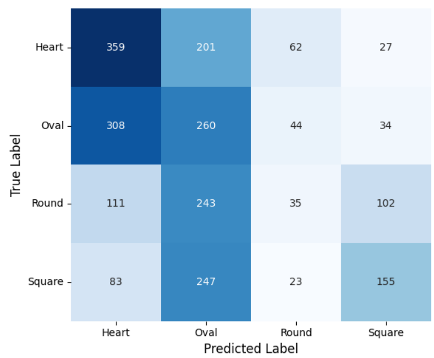

Fig. 3. Confusion matrix for the threshold-based classifier.

B. Machine-Learning Classification Results

To address the limitations of handcrafted rules, two machine-learning models were trained using the same set of geometric ratios: Random Forest and Gradient Boosting. Both models substantially outperformed the threshold-based approach. Random Forest achieved an accuracy of 50.00%, and Gradient Boosting reached 48.25%.

Even though these results represent a significant improvement, the moderate accuracy values suggest that geometric ratios alone are still limited in capturing the complexity of face-shape differences.

Figures 4 and 5 present the confusion matrices for the learning-based models. Compared to the threshold-based classifier, both models exhibit more balanced distributions across predicted categories, indicating a better ability to generalize facial proportion patterns.

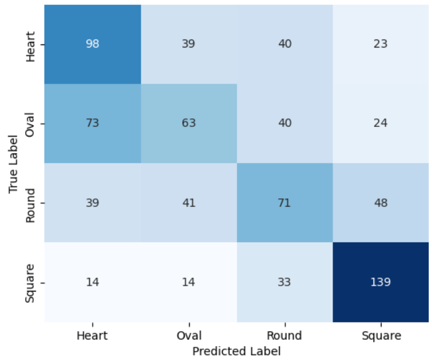

Fig. 4. Confusion matrix for the Random Forest classifier.

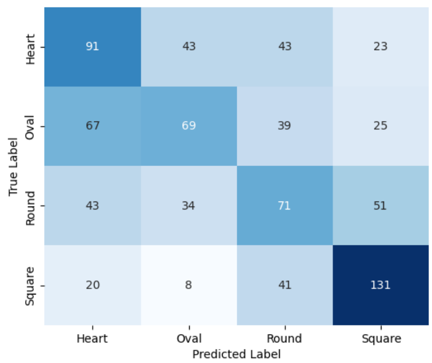

Fig. 5. Confusion matrix for the Gradient Boosting classifier.

C. Comparison Across Methods

Both machine-learning models more than doubled the accuracy obtained with the threshold-based classifier. Their confusion matrices show improved class balance and less severe prediction collapse. However, the remaining errors across all methods emphasize that relying solely on low-dimensional geometric ratios may not provide sufficient expressive power for robust face-shape classification. These results motivate future work incorporating higher-dimensional landmark embeddings or deep-learning representations to improve predictive performance.

V. Conclusions

This work presented a face-shape classification pipeline combining landmark-based geometric analysis and machine-learning models. Although simple geometric thresholds provided limited accuracy, the learning-based methods significantly improved performance, demonstrating that geometric ratios contain relevant but insufficient information for fine-grained classification. Future work may incorporate higher-dimensional landmark embeddings or deep-learning models to capture more complex morphological patterns and further enhance classification reliability.

References

[1] S. Hangaragi, T. Singh *et al.*, "Face detection and recognition using face mesh and deep neural network," *Procedia Computer Science*, vol. 218, pp. 741–749, 2023.

[2] W. Sunhem and K. Pasupa, "An approach to face shape classification for hairstyle recommendation," in *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, 2016, pp. 390–394.

[3] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.

[4] M. Hossam, A. A. Afify, M. Rady, M. Nabil, K. Moussa, R. Yousri, and M. S. Darweesh, "A comparative study of different face shape classification techniques," in *2021 International Conference on Electronic Engineering (ICEEM)*. IEEE, 2021, pp. 1–6.

[5] A. Chaudhary, S. Kolhe, and R. Kamal, "An improved random forest classifier for multi-class classification," *Information Processing in Agriculture*, vol. 3, no. 4, pp. 215–222, 2016.

[6] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review*, vol. 54, no. 3, pp. 1937–1967, 2021.