Hindawi
Complexity
Volume 2022, Article ID 6709707, 14 pages
https://doi.org/10.1155/2022/6709707

Research Article

Caricature Face Photo Facial Attribute Similarity Generator

Muhammad Irfan Khan, Muhammad Kashif Hanif, and Ramzan Talib

Department of Computer Science, Government College University, Faisalabad, Pakistan

Correspondence should be addressed to Muhammad Kashif Hanif; mkashifhanif@gcuf.edu.pk

Received 27 May 2021; Revised 10 December 2021; Accepted 28 January 2022; Published 22 February 2022

Academic Editor: Muhammad Ahmad

Caricatures can help to understand the perception of a face. The prominent facial feature of a subject can be exaggerated, so the subject can be easily identified by humans. Recently, significant progress has been made to face detection and recognition from images. However, the matching of caricature with photographs is a difficult task. This is due to exaggerated features, representation of modalities, and different styles adopted by artists. This study proposed a cross-domain qualitative feature-based approach to match caricature with a mugshot. The proposed approach uses Haar-like features for the detection of the face and other facial attributes. A point distribution measure is used to locate the exaggerated features. Furthermore, the ratio between different facial features was computed using different vertical and horizontal distances. These ratios were used to calculate the difference vector which is used as input to different machine and deep learning models. In order to attain better performance, stratified k-fold cross-validation with hyperparameter tuning is used. Convolution neural network-based implementation outperformed the machine learning-based models.

1. Introduction

A picture is worth a thousand words—a truth that is derived from experience [1]. It refers that an image can convey complex and even multiple ideas in a more effective way as compared to verbal description. Cartoons and caricatures are designed by artists to represent a short story or theme within an image. A cartoon is a nonrealistic or sometimes semirealistic style to interpret or visually explain some concept. According to Kleeman, cartoon is enjoyable for someone if he understands the cartoonist's viewpoint [2]. James Gillray has been known as the founder of political cartoons [3]. Apparently, caricatures and cartoons are entirely similar in nature. Caricatures and cartoons can be sketched either by some humans or by a computer. The element of verisimilitude makes a caricature different from the cartoon. Caricatures are always of real persons, while the cartoon can be fabricated for unrealistic persons. Caricature is an image of a person (politician or someone famous in the public) made by an excessive depiction of some characteristics of the person. Caricatures, in nature, capture the physical traits of some person which are exaggerated for humorous effects. For example, if the ears of a person are much prominent than an average one, then ears will be portrayed much larger than common.

The human brain can identify the person depicted in a caricature more quickly as the exaggerated features act as an eye-catcher and thus help to identify the person in a short time (Figure 1). According to [5], well-designed caricatures are more easy to recognize than perfect portraits.

Researchers have designed different applications and algorithms for face detection and recognition [6]. Most of these applications work for heterogeneous face recognition. However, these techniques are not appropriate for recognizing a person from caricatures.

This work tackles the problem of matching a caricature to a mugshot using different machine and deep learning algorithms. In order to match caricature to a photograph, qualitative facial attributes are defined. The qualitative facial features such as forehead height and width and nose height and width were used to encode the physical appearance of the face. These qualitative features help to determine whether a face is a mean face or digressed from a mean face. We proposed statistical learning techniques to measure the

Figure 1: Examples: photos and corresponding caricatures. For each subject, the first column contains a photo, and the next 3 columns contain caricatures by different caricaturists [4].

weights of these features. These features were used in machine and deep learning algorithms to detect and recognize a face.

In this work, different approaches were discussed to match caricature with a mugshot. The contributions of this study are described as follows:

(i) Some features are more explanatory for caricatures when compared with photographs. For this reason, analytical representation of facial components of caricature and photographs was used.

(ii) We employed Haar-like features to tackle the challenge of facial feature extraction from exaggerated artistic work in caricature.

(iii) In order to detect facial landmarks, different horizontal and vertical distances were computed using the Euclidean distance. Moreover, we devised different thresholds for different facial attributes to detect face shape and to incorporate maximum features for better performance.

(iv) Attribute-based proportionality technique was used to minimize the cross-domain differences.

(v) A difference vector was computed based on qualitative features. Then different machine and deep learning-based algorithms were employed to characterize the performance.

The rest of the paper is organized into different sections. Section 2 describes the heterogeneous face recognition and role of facial attributes. Section 3 discusses related work. Section 4 describes the proposed technique used in this work. Results and evaluations of the proposed method are elaborated in Section 5. Section 6 gives the conclusion of this study.

2. Heterogeneous Face Recognition

Heterogeneous face recognition (HFR) is a known paradigm for face recognition and matching of different modalities. One of the major applications of HFR is sketch-based face recognition (SBFR) which deals with matching facial sketches to photographs. SBFR can be classified on the basis of how the sketches are produced. There exist four widely used categories of sketches (Figure 2):

(i) Viewed sketches: mugshots are referred to as artists (Figure 2(a))

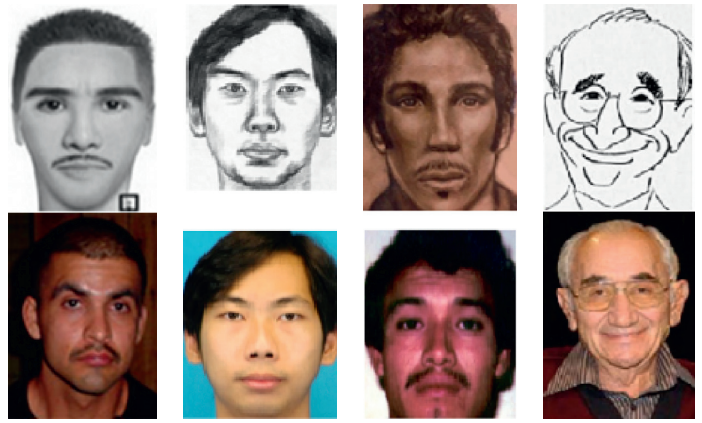
(ii) Forensic sketches: hand-drawn sketches on the basis of witnesses (Figure 2(b))

(iii) Composite sketches: made by experts by using specific software (Figure 2(c))

(iv) Caricature sketches: sketches with exaggerated facial features (Figure 2(d))

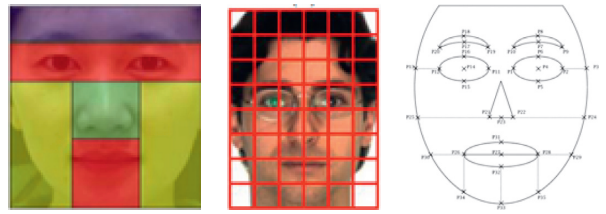
The eigen-transformation algorithm which was proposed by Tang and Wang [11] is the foundation of synthesis-based HFR techniques. In the HFR system, the first step is to characterize face images in different modalities. The most common representations of the face are holistic, patch-based, component-based, and analytics (Figure 3).

Each face part (e.g., mouth, nose, eyes, and lips) can be represented independently using component-based representations [14] (Figure 3(a)). This can help to separately measure the features of each component in the matching process. Match score among two face images can be produced by some component-fusion scheme. The capabilities of both linear and nonlinear misalignment across modalities



(a) (b) (c) (d)

Figure 2: The first row is of sketches and the second row is of mugshot photos. (a) A composite sketch created using the software FACES [7, 8]. (b) Viewed sketch [9]. (c) Forensic sketch [9]. (d) Caricature [10].



(a) (b) (c)

Figure 3: (a) Five facial components corresponding to the eye, eyebrow, cheeks, nose, and mouth. (b) Patch-based holistic representation [12]. (c) Analytical representation with fiducial points [13].

can be improved when components are detected correctly and matched [14].

Face image can also be represented by a single vector using global holistic representations [15]. The appearance of each image is encoded in patches with a feature vector for each patch in patch-based holistic representations (Figure 3(b)). There exist a variety of approaches to use these patches, for example, learning a classifier per patch [16] or concatenation as a large feature vector [11]. This approach can encode all information of available appearance. However, a high-dimensional feature vector can have sensitivity for expression and alignment variations which can result in overfitting [17].

In analytical representations, the face is modeled geometrically by detecting facial components and fiducial points on the face [18] (Figure 3(c)). This representation is relatively invariant to modality when fitting a model to a face in different modalities. However, it can require human involvement to avoid ineffectiveness in face-based model fitting. Moreover, it does not support facial expressions [18].

3. Related Work

This section presents the previous work related to feature-based heterogeneous face recognition for caricature and sketch. In HFR, feature-based approaches concentrate on developing a feature descriptor for the images. These feature descriptors are unvarying to the modality but varying to the identity of the person. The popular image feature descriptors are Gabor transform, scale-invariant feature transform (SIFT) [11], local binary pattern (LBP) [19], and histogram of averaged oriented gradients (HAOG) [20]. Sketch and photographs are matched directly after they are encoded using one of these descriptors. Table 1 summarizes the recognition approaches used in feature-based HFR.

SIFT features offer a compact vector representation of the image. Klare et al. [11] proposed an invariant SIFT feature-based method to match sketches and photographs. They modeled SIFT feature vectors from the mugshot images and concatenated them jointly for sketches and mugshot images. Then, Euclidean distances are estimated between concatenated SIFT feature vectors of both sketch and mugshot images for nearest neighbor (NN) matching.

Bhatt et al. [23] proposed a method based on extended uniform circular LBP descriptors. They also employed a weight optimization technique based on genetic algorithm (GA) [26] to seek optimum weights for each facial patch. Lastly, NN matching is executed by using the chi-square distance measure.

A self-similarity descriptor was proposed by Khan et al. [21]. Features were extracted individually from local regions of photographs and sketches. A small image patch is correlated within its larger neighborhood to obtain the self-similarity features. Self-similarity reduces the modality gap

Table 1: Feature-based HFR matching methods.

Features	Recognition approach	Publication
SIFT	NN	[11]
Self-similarity	NN	[21]
Gabor shape	NN, chi-square	[22]
HAOG	NN, chi-square	[20]
EUCLBP	Weighted chi-square	[23]
LRBP	NN, PMK, chi-square	[24]
CITE	PCA + LDA	[16]
Geometric features	K-NN	[25]

since it remains comparatively invariant to the sketch-photograph modality variation.

Galoogahi et al. [24] proposed Local Radon Binary Pattern (LRBP) as a new face descriptor. This descriptor was helpful to directly match sketches and mugshots. In this framework, the mugshots are transfigured into Radon space. Afterward, these transformed images are encoded by LBP. As a final step, LRBP is computed by concatenating the histogram of local LBPs. A distance measurement based on pyramid match kernel (PMK) [27] is used to perform matching.

Another face descriptor was introduced by Zhang et al. [16] on the basis of coupled information-theoretic encoding. This descriptor uniquely captures discriminative local facial structures. Thus, a coupled encoding was obtained through an information-theoretic projection tree. Klare et al. [11] combined their SIFT descriptor with a common representation space strategy which was projection-based. This improvement was made on the assumption that even though direct comparison of sketches and mugshots is not possible, the distribution of interface similarities will be comparable within the mugshot and sketch domain. Consequently, for each sketch and mugshot, re-encoding is performed to obtain a vector of their Euclidean distances for the training set of sketches and mugshots accordingly. This common representation acts as the invariant to modality.

The caricature recognition task is similar to forensic sketch recognition. Besides the modality shift challenge, they hold either incomplete or imprecise information due to the judgment based on witness' personal feelings and opinions, and shortcoming of memory. A system for automatically matching sketches to photographs was proposed by Uhl et al. [28]. They geometrically standardized the sketch and mugshot to assist comparison after extracting the facial features from sketches and mugshots. As the final step, eigen analysis was performed for matching. Although their method was outmoded as compared to modern methods, they attracted researchers towards forensic sketch and caricature-based face recognition problems. A study was carried out by Klare et al. [8] in which they introduced an approach that utilized projection-based and feature-based contributions. They presented a framework named local feature-based discriminant analysis (LFDA). In LFDA, they independently represent both photos and sketches using SIFT and multiscale local binary patterns (MLBPs).

The algorithm proposed by Bhatt et al. [19] combines the projection- and feature-based contributions to enhance recognition achievement. They encoded structural information in local facial regions by using multiscale circular Webber's local descriptor.

Generally, it is considered that the caricature recognition problem can be solved by sketch recognition methods. During earlier research, a semantic face graph was proposed to match facial mugshots that were converted into caricatures [29]. A photograph is transformed into a sketch (or vice versa) so that cross-modal differences could be eliminated. Different face recognition methods can be employed in this transformation (photograph to sketch or sketch to photograph) [30]. The cross-model differences are comparatively more than that of view-based sketches (sketches drawn from a photograph). For this purpose, the photographs and sketches are encoded into a common space MLBP [31] and SIFT [32] descriptors. Weights are allocated to facial regions by using a multiscale circular Weber's descriptor [19]. To minimize cross-domain gaps, 68 facial attributes were proposed [33]. However, automatic extraction of these facial features was left unsolved. A midlevel attribute representation was used to define a method for cross-modality matching [34].

For caricature and photograph matching, there are some commonly used datasets for benchmarking (Table 2). Each dataset consists of pairs of photographs and caricatures/sketches. However, these datasets differ based on viewed sketches, drawn by an artist, and the sketches drawn by using the software.

The existing work seems to be tedious due to the extraction of facial attribute features. The attribute features used are labeled by humans [33]. Based on these features, support vector machines (SVMs), multiple kernel learning (MKL), and logistic regression (LR) were applied to estimate the similarity level of a photograph and a caricature. A method to extract the facial attribute features from a photograph was proposed by Klare et al. [9]. However, manual work was done to identify the attribute features of caricatures. They used a genetic algorithm to find the weights of these attributes.

In most of the existing work, facial attributes were labeled manually for caricatures. In this study, qualitative feature extraction is performed automatically. We tried to minimize the cross-domain differences by using attribute-based proportionality. Furthermore, exaggerated features are difficult to handle in Haar-like features. Therefore, we employed qualitative feature matching using machine and deep learning techniques.

4. Methodology

This section presents the methodology adopted to match caricature with a mugshot. Face detection is the most fundamental and essential part of the face recognition process. Caricature face detection is distinct from that of face detection from mugshots. The reason is that caricatures do not preserve facial features to a massive scope. Moreover, the variations may occur in caricatures regarding the artistic style. Therefore, facial feature extraction and face detection

Table 2: Existing datasets for caricatures.

	Dataset information	
	Subjects	Images
Klare et al. [33]	196	392
Abaci and Akgul [9]	200	400
Mishra [4]	100	Caricature: 8,928, face: 1,000

in caricature are much complicated as that of mugshots or other photos.

Different techniques can be employed due to variations in the properties of caricatures and photographs. These variations are differences in resolution, color, and skin textures. Figure 4 shows the proposed methodology. The most basic step is to identify the qualitative features of photos and caricatures. These qualitative features are evaluated to check the matches. Higher values signify the correct matching of photographs and caricatures. Numerical encoding of features is comparatively different due to exaggerated features in caricature. For example, a wider nose is exaggerated by the artist to make it prominent. Thus, caricatures can grab and exaggerate the prominent features of a person/face while the unimportant features are discarded. The artists outline the caricature by drawing face outlines such as lips, eyes, and nose in physically exact locations with few exaggerations to make it funny but identifiable. Furthermore, some other basic attributes such as hairstyle (or some special cap/turban), beard, eye-glasses, and mole are also drawn to distinguish the caricature. For this reason, we used some qualitative facial features.

Caricature and photographs can have noise or blur images. Preprocessing is performed in order to match caricatures with photographs. This step removes noise, converts photos to grayscale, and normalizes the size with respect to eye and mouth coordinates. Then, the Haar cascade classifier was employed to detect faces from caricatures and photographs using the Haar features [35]. The reason to use this algorithm is the high detection rate and fast processing. This technique integrates different classifiers which can eliminate nonface regions within an image. Moreover, the AdaBoost algorithm is used in this approach to take important features.

First, fewer explanatory features in face appearance [36] are observed. These less explanatory features are useful for caricature recognition. These features include head shapes (i.e., oval, rectangle, circle, square, heart, or triangle) (Figure 5). In addition to head shape features, some facial attributes such as nose (Figure 6) and eye shapes (Figure 7) are used. We focused on extracting eyes, nose, lips, and chin from the image (Figure 8). Moreover, we calculated the sum of pixels in a feature window by integral image concept to avoid the summing up individually. More training data are required to extract the facial features from caricatures (Figure 9).

Facial landmarks play an important role to determine facial features. The distances between these landmarks are the key information that must be used effectively. The horizontal and vertical distances (Figures 10 and 11) between different facial attributes are important in caricature recognition. The reason is these features are also exaggerated by an artist to create a caricature.

To determine the shape of the face, we marked horizontal distances between some facial landmarks (Figure 12) which are H_1, H_2, H_3, H_4, H_5 and vertical distances from the respective horizontal line to the Y-coordinate of the deepest landmark present at the chin as V_1, V_2, V_3, V_4, V_5.

Horizontal distances H_1, H_2, H_3, H_4, H_5 are computed using Euclidean distance as follows:

$$H_i = \sqrt{(x_{12-i})^2 - (x_i)^2}, \quad 1 \le i \le 5. \tag{1}$$

Vertical distances V_1, V_2, V_3, V_4, V_5 are calculated using the following equation:

$$V_j = \sqrt{(y_6)^2 - (x_j)^2}, \quad 2 \le j \le 5. \tag{2}$$

Index of minimum and maximum horizontal distance $N_{\min}$ and $N_{\max}$ from H_2, H_3, H_4, H_5 is calculated using equation (3) and equation (4), respectively.

$$N_{\min} = \arg_n \min\left(\left(H_2 | H_3 | H_4 | H_5\right) \le 0.95 H_1\right)\right), \tag{3}$$

$$N_{\max} = \arg_n \max\left(\left(H_2 | H_3 | H_4 | H_5\right) \le 0.95 H_1\right)\right). \tag{4}$$

$N_{\max}$ is used to find maximum horizontal distance and relevant vertical distance using equation (5) and equation (6):

$$H_{\max} = H_{N_{\max}}, \tag{5}$$

$$V_{H_{\max}} = V_{N_{\max}}. \tag{6}$$

Then two ratios r and t are computed the following equations:

$$r = \frac{V_{H_{\max}}}{H_{\max}}, \tag{7}$$

$$t = \frac{V_5}{V_2}. \tag{8}$$

Next, a threshold value of 0.35 is used to determine the category of the face. If t is less than equal to 0.5, then the face shape is square; otherwise, it is rectangle.

if $r \le 0.35$ then
if $t \le 0.5$ then
Face is Square
else
Face is Rectangle
end if
end if

If the ratio $r \ge 0.35$, then θ_1 and θ_2 are computed using the following equation:

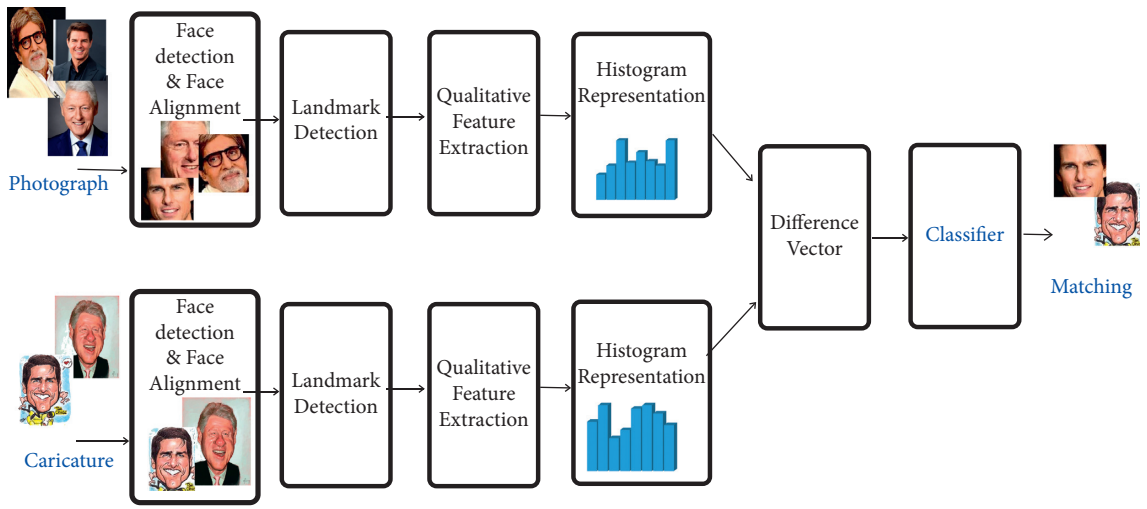


Figure 4: Overview of the proposed model for caricature identification.

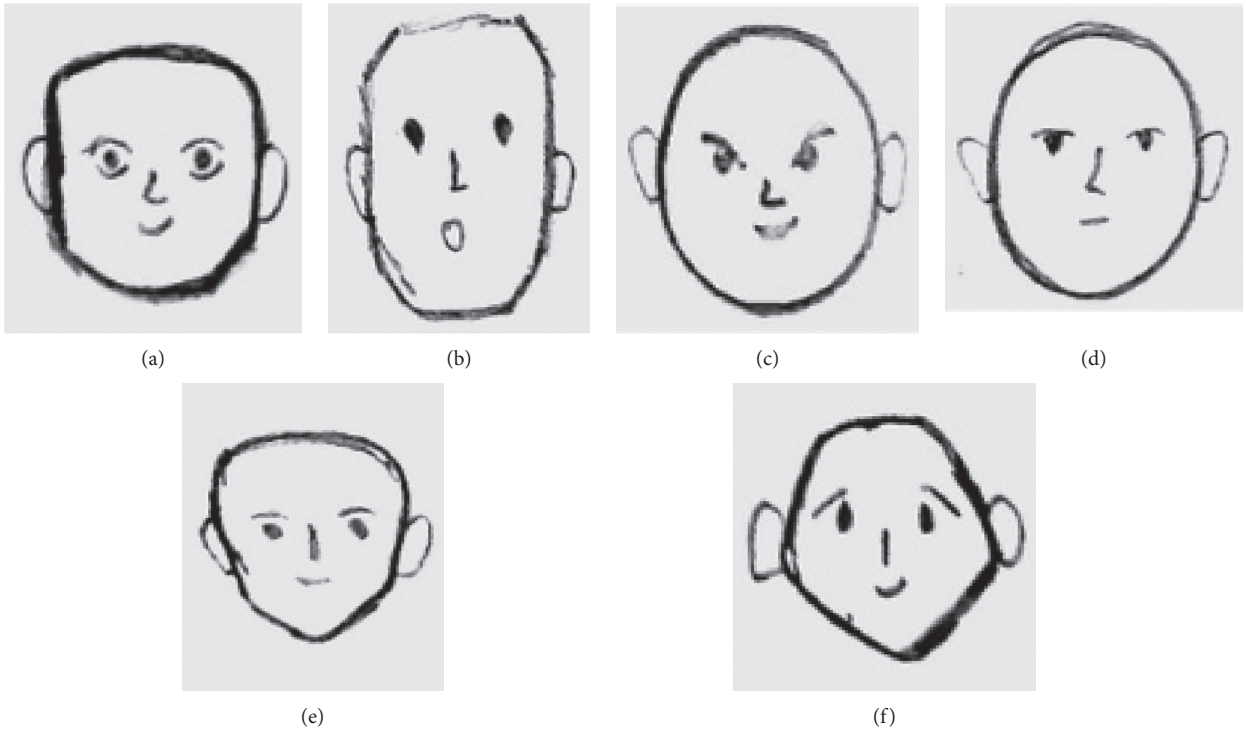


Figure 5: Features: generally less explanatory for photos but much explanatory for caricatures. (a) Square face, (b) rectangle face, (c) circle face, (d) oval face, (e) heart face, and (f) triangle face.

$$\theta_1 = \arctan\left(\frac{y_3 - y_4}{x_3 - x_4}\right),$$
$$\theta_2 = \arctan\left(\frac{y_4 - y_5}{x_4 - x_5}\right). \quad (9)$$

Then r is recalculated as

$$r = \frac{|\theta_1 - \theta_2|}{65}, \quad (10)$$

and face shape is recognized by

if $r \leq 0.35$ then
if $t \leq 0.5$ then
Face is Circle
Else
Face is Oval
end if
else

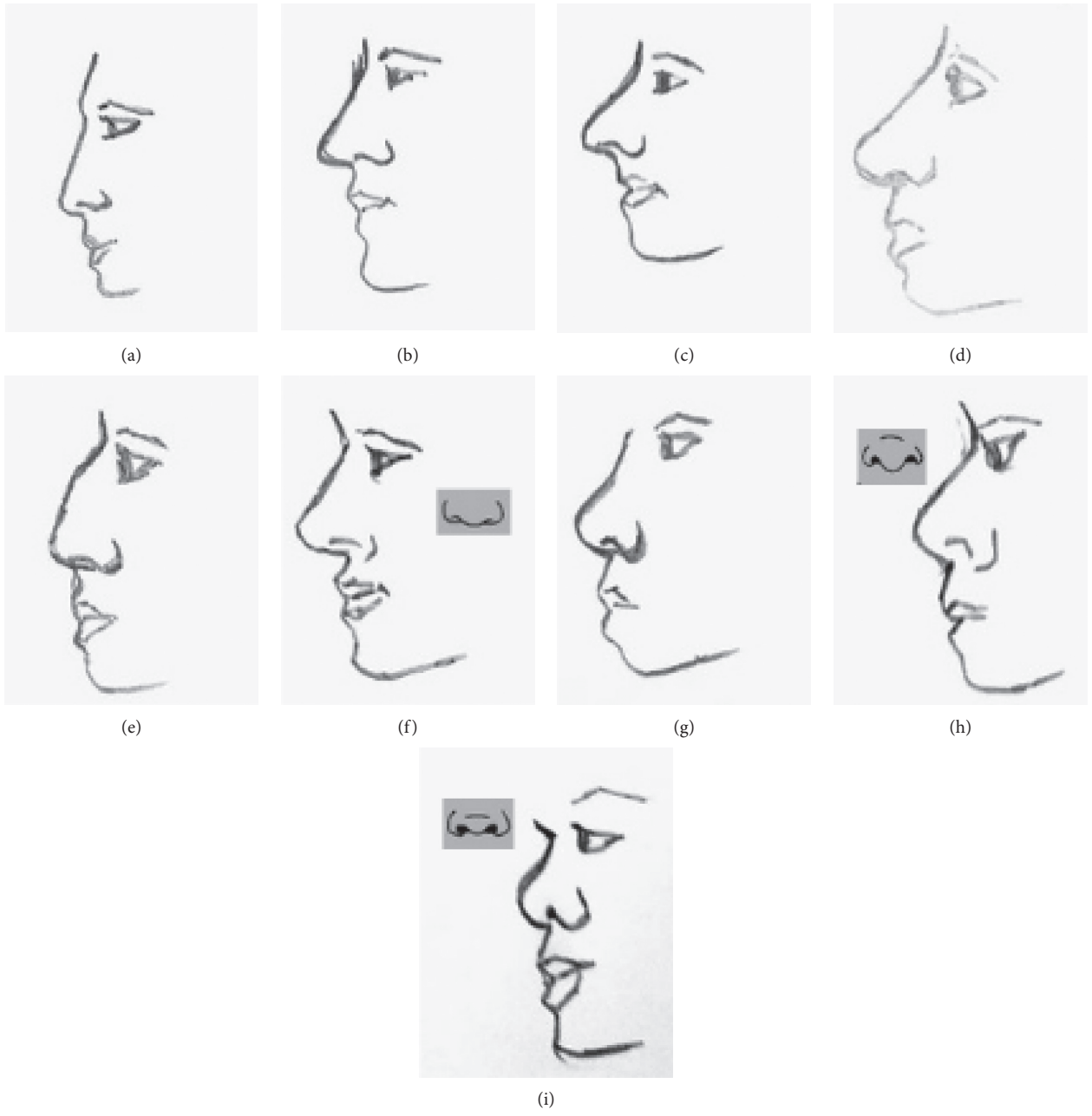


Figure 6: Nose shapes: (a) Grecian nose (drops straight down from the forehead), (b) Roman nose (slightly aquiline), (c) Aquiline nose (eagle-like convex), (d) droopy nose (tip very low, an effect of aging), (e) hooked nose (broken profile), (f) button nose (rounded and small, tip turns up so slightly that nostrils are not visible), (g) upturned nose (concave), (h) snub nose (aka blunt, short, and upturned, mostly found in Asians), and (i) funnel nose (African nose, nostrils pass over to the bridge).

if $t \leq 0.5$ then
Face is Heart
else
Face is Triangle
end if
end if

Furthermore, a difference vector is calculated for each possible pair of caricatures and photographs in the training set. The difference vector is marked as +1 and −1 for true and false match of caricature and photograph, respectively. Suppose $\{(Z_i, t_i), Z_i \in \mathbb{R}, t_i \in \{-1, +1\}, \quad i = 1, 2, \ldots, n\}$ are the n-pair of difference vectors. t_i is +1 if Z_i is a difference vector for a true match of caricature and photo. t_i is set to −1 for false match. Thus, for n-subjects in the training set, positive samples for true matches are n, whereas negative samples for false matches are $n^2 - n$.

For improved qualitative feature matching mechanisms, machine learning approaches for the selection and weighting of feature subsets are employed. In this study, we compared the performance of LR, SVM, MKL, and convolutional

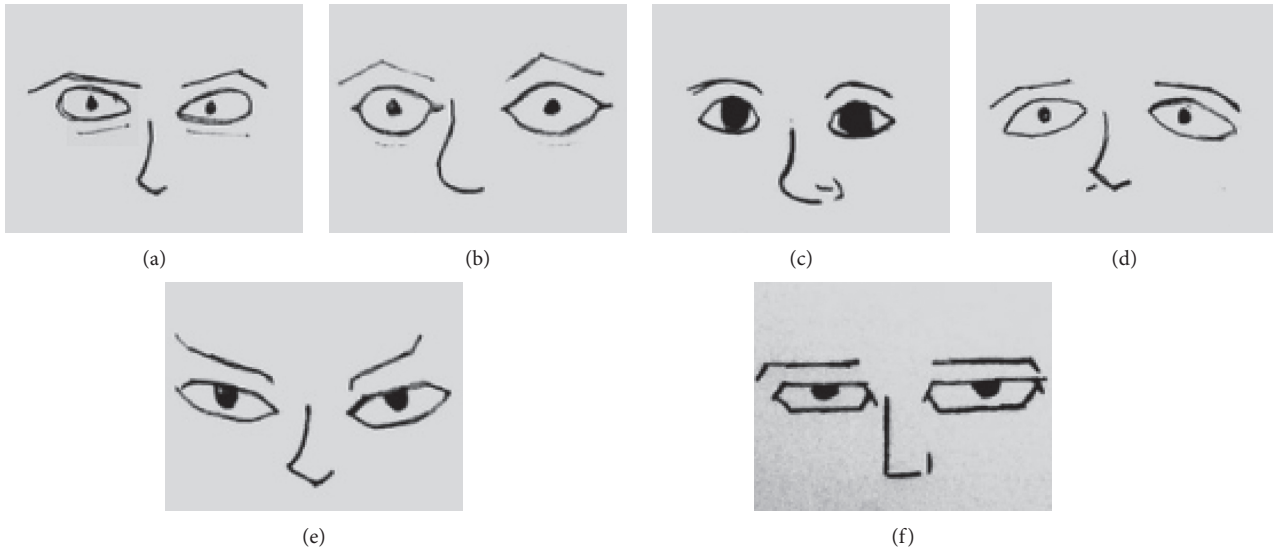


Figure 7: Eye shapes: (a) almond eye, (b) round-shaped eye, (c) small eye, (d) downturned eye, (e) upturned eye, and (f) deep set eye.

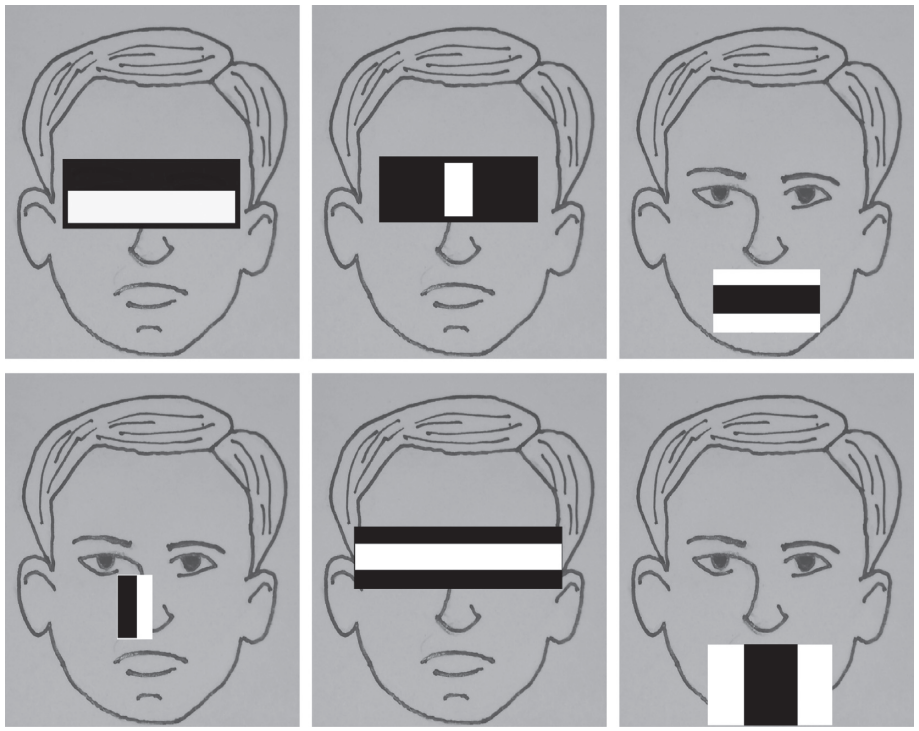


Figure 8: Haar features for facial attributes.

neural network (CNN) for matching. LR searches for a function that outlines the difference vectors to numerical labeling (here these are +1 or −1). A similarity score is obtained as an output. Similarity value is computed as

$$f(Z) = X\alpha - \log(\exp(Z\alpha) + 1), \tag{11}$$

where Z is the difference vector between a caricature and a photograph. The drawback of the LR method is it works only for the linear dependency of features. For this reason, MKL and SVM are employed to work with nonlinear dependencies [37].

For n training images, a set of base kernels is $\{F_i \in \mathbb{R}^{n\times n}, \quad i = 1, 2, \ldots, 25\}$. To combine these base kernels, the coefficient $q = (q_1, q_2, \ldots, q_r)^T \in \mathbb{R}_+^r$ is used. The kernel matrix is $F(q) = \sum_{i=1}^{r} q_i F_i$. Convex-concave optimization of the MKL dual formulation is used to get the coefficient vector q [38]. MKL is used with the nearest neighbour. This helped to obtain the weighted differences for each feature vector. SVM algorithm is used with a single kernel utilizing all feature components at the same time [39].

Moreover, this study also employs CNN for caricature and photograph matching. Input images are resized and

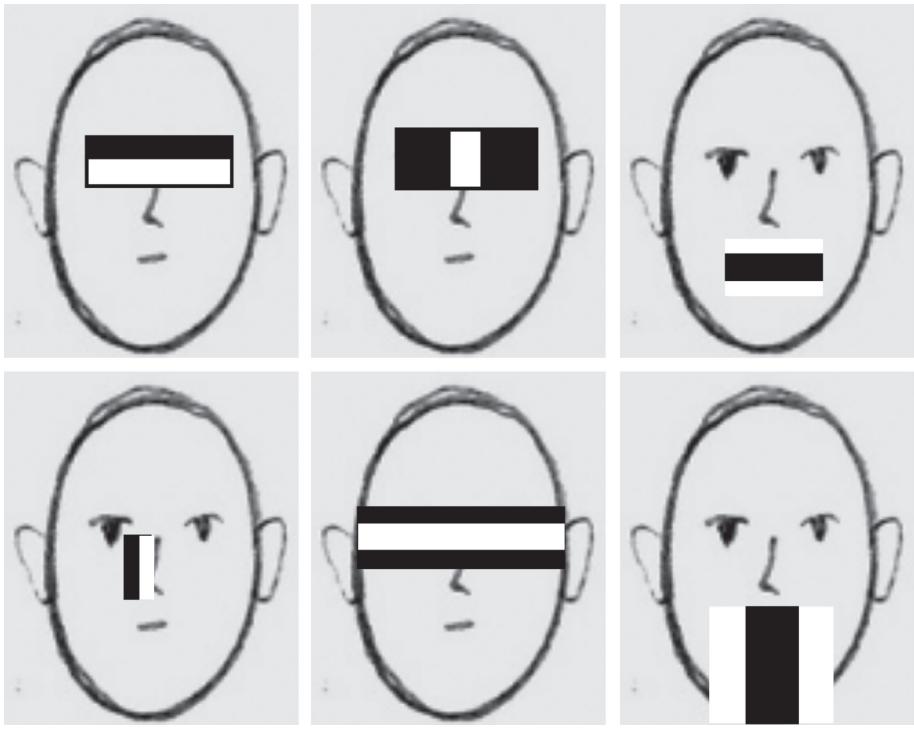


Figure 9: Haar features for facial attributes in caricatures.

padded to size 224 × 224. Then these images are converted to grayscale. These transformed images are used as input to CNN. The proposed CNN architecture consists of ten convolutional layers, five pooling layers, and one fully connected layer. Max operator is applied in four pooling layers and the last max-pooling layer uses average. The approximation of a large filter is obtained by applying several small filters. Also, the number of parameters is turned down by eliminating the redundancy of fully connected layers.

5. Results and Discussion

This section discusses the results of techniques used to match caricature with a photograph. Google Colab is used for the implementation of machine and deep learning models. Google Colab is a free online cloud-based Jupyter notebook environment having high computational power. We used the IIIT-CFW database [4]. There are 8929 cartoon faces and 1000 genuine faces of 100 renowned personalities. For training, 80% of the dataset is used and the rest of the dataset is used for testing.

We employed stratified 10-fold cross-validation and hyperparameter tuning to find the best parameters for each approach. Stratified k-fold cross-validation helps to measure the performance by splitting the dataset into k subsets. One of these subsets is taken as a testing subset and others as training subsets. This procedure is iterated k times for different subsets. Stratified k fold cross-validation divides the data such that the proportions between classes are the same in each fold [40]. The machine and deep learning models were retrained for the best parameters obtained using hyperparameter tuning.

Figures 13 and 14 show the result of the proposed model using precision, recall, F1-scores, and normalized confusion matrix of five celebrities selected from the dataset. The proposed model has a high value of precision and recall (Figure 13). This means the proposed model has a low false-positive rate and a low false-negative rate. Diagonal values in the normalized confusion matrix also validate the performance of the proposed model (Figure 14).

Moreover, receiver operating characteristic (ROC) and cumulative match characteristic (CMC) analysis are used to evaluate the performance. The ROC analysis is made by true-positive rate (TPR) versus false-positive rate (FPR). TPR and FPR calculations are given in equations (12) and (13). The accuracy is computed using equation (14).

$$\text{TPR} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \times 100, \tag{12}$$

$$\text{FPR} = \frac{\text{true negative}}{\text{true positive} + \text{false negative}} \times 100, \tag{13}$$

$$\text{accuracy} = \frac{\text{TPR} + \text{FPR}}{2} \times 100. \tag{14}$$

The results of TPR at fixed FPR of 1% and 10% are listed in Table 3.

The CMC curve shows the caricature recognition accuracy (Figure 15). CMC evaluates the frequency that a caricature is

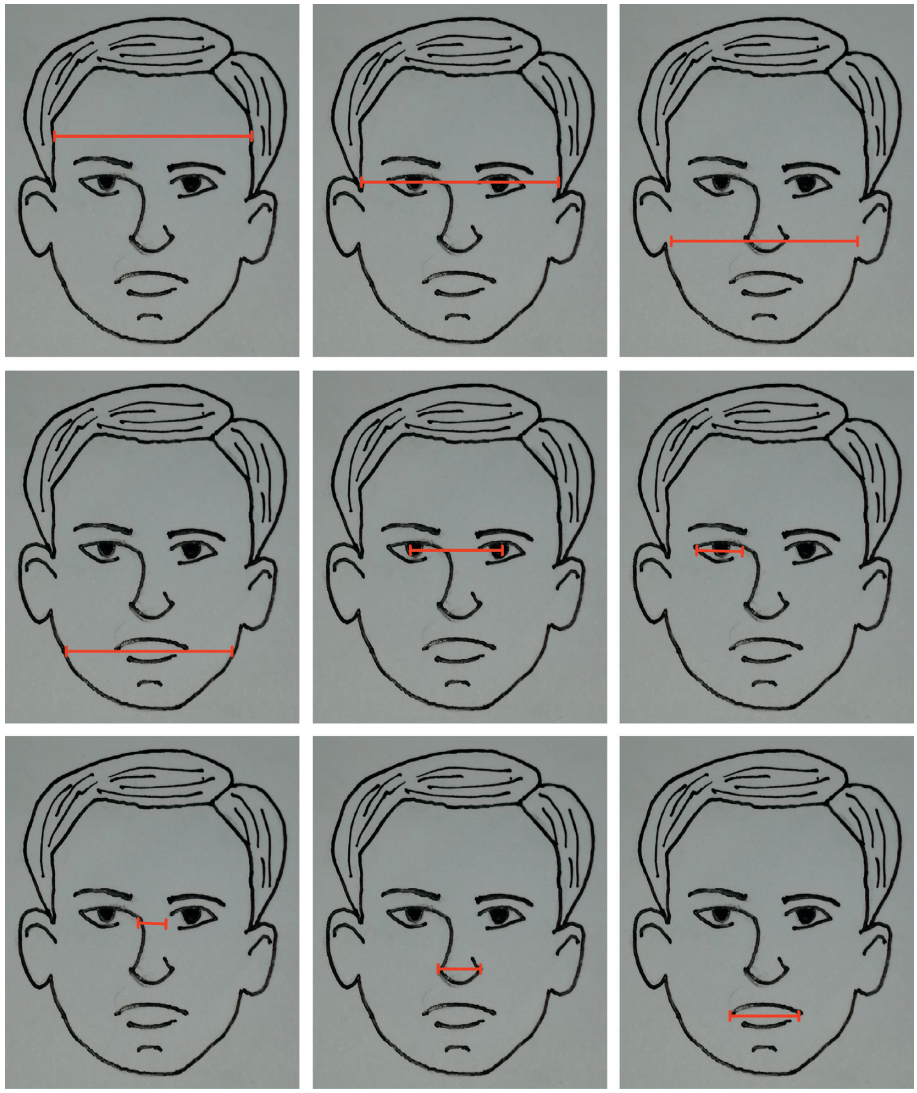


Figure 10: Horizontal distances between landmarks for basic facial attributes.

matched with the same identity when it is searched in a dataset of photos. The ranks are plotted at the X-axis and the frequency percentage is plotted at the Y-axis. The percentage of times that at least one out of the top n matches in the dataset is the same caricature is plotted as the frequency at Rank n. The Rank 1 and Rank 10 scores are listed in Table 4. Our proposed approach attained approximately 78% accuracy using CNN. The existing study attained 74% accuracy [33]. The reason for improved performance is using proportionality for facial features. However, the existing study used different weights for facial features.

Testing is performed by training the algorithm only on photographs (without caricatures). It is observed that when knowledge is transferred from the mugshot domain to the caricature domain, the results are improved a certain percentage. We can infer the importance of different qualitative features with the help of vector estimation. We can find the important features using the assigned weights to these qualitative features.

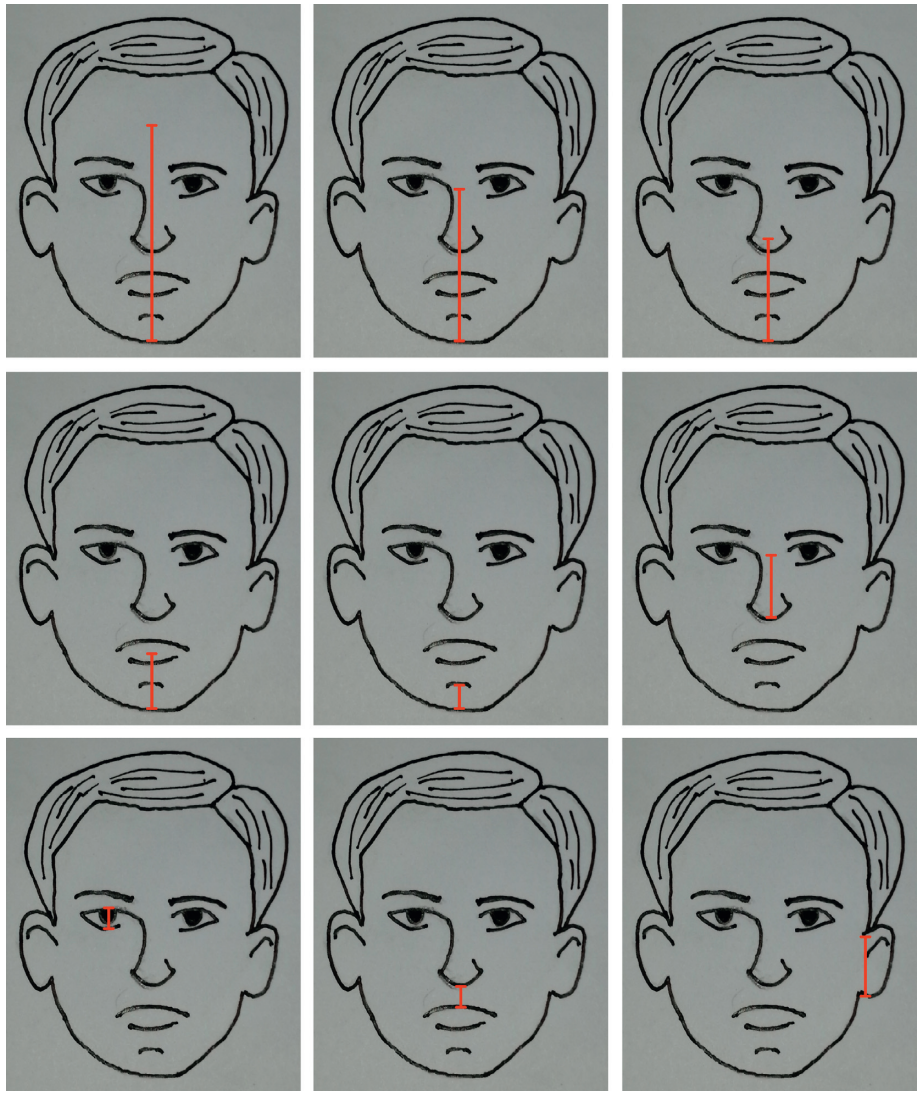


Figure 11: Vertical distances between landmarks for basic facial attributes.

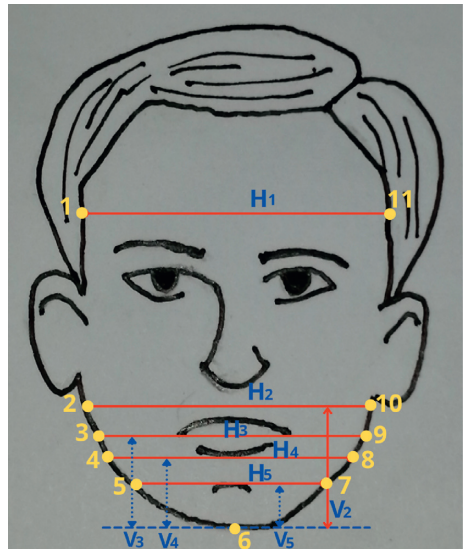


Figure 12: Basic distances between facial landmarks to determine the face shape.

	precision	recall	f1-score
Obama	0.91	0.95	0.93
Hillary	0.93	0.87	0.90
Amitab	0.89	0.94	0.91
Clinton	0.95	0.95	0.95
Tom Cruise	1.00	0.93	0.96
accuracy			0.93
macro avg	0.94	0.93	0.93
weighted avg	0.93	0.93	0.93

Figure 13: Confusion matrix of caricature recognition results.

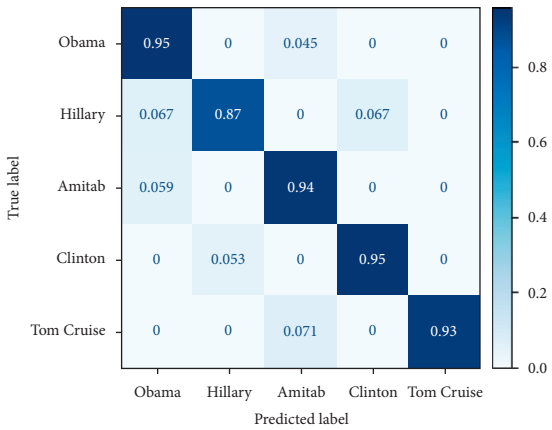


Figure 14: Normalized confusion matrix.

Table 3: Performance analysis of different classifiers.

Method	TPR at FPR = 1%	TPR at FPR = 10%
LR	13.1 ± 2.1	54.7 ± 3.3
SVM	13.9 ± 1.9	58.6 ± 4.3
MKL	9.4 ± 3.8	45.4 ± 3.9
CNN	24.6 ± 2.6	73.4 ± 4.7

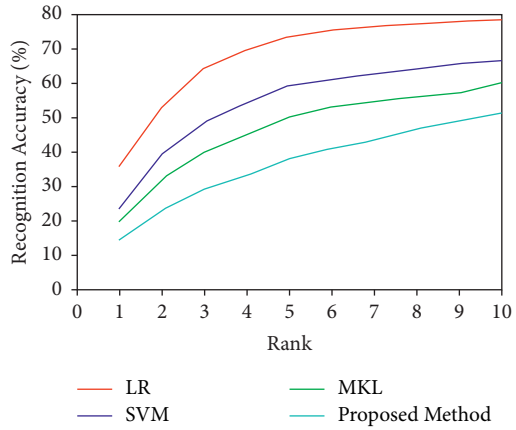


Figure 15: Successful caricature to photo-matching rates (Rank 1 to Rank 10).

Table 4: Average scores of caricature recognition accuracy (%) for different classifiers.

Method	Rank 1	Rank 10
LR	19.7 ± 2.1	59.8 ± 3.3
SVM	23.3 ± 4.6	66.4 ± 2.9
MKL	14.1 ± 2.9	51.1 ± 3.7
CNN	35.8 ± 2.7	78.3 ± 3.3

6. Conclusion

Caricature face detection and feature extraction required extensive effort because of misalignment problems due to exaggeration of facial features. This study proposes a cross-domain facial qualitative feature matching of caricature with photographs. Haar was employed to detect and extract the facial features from caricatures and mugshots. Euclidean distance was used to compute vertical and horizontal distances to calculate the ratio between different facial attributes. A difference vector based on qualitative features was designed which is used as input for different machine and deep learning algorithms. The proposed approach using CNN performed better when compared with other techniques and attained approximately 78% accuracy. In the future, we are interested to identify and match pose variant caricatures using different machine and deep learning techniques.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

[1] W. Gardner, *Speakers give sound advice*, Syracuse Post Standard, New York, NY, USA, 1911.

[2] G. Kleeman, "Not just for fun: using cartoons to investigate geographical issues," *New Zealand Geographer*, vol. 62, no. 2, pp. 144–151, 2006.

[3] M. Rowson, "Satire, sewers and statesmen: why james gillray was king of the cartoon," *The Guardian*, vol. 16, 2015.

[4] A. Mishra, S. N. Rai, A. Mishra, and C. V. Jawahar, "Iiit-cfw: a benchmark database of cartoon faces in the wild," in *Proceedings of the European Conference on Computer Vision*, pp. 35–47, Springer, Amsterdam, Netherlands, October 2016.

[5] R. Mauro and M. Kubovy, "Caricature and face recognition," *Memory & Cognition*, vol. 20, no. 4, pp. 433–440, 1992.

[6] P. J. Grother, P. J. Grother, P. J. Phillips, and G. W. Quinn, *Report on the evaluation of 2D still-image face recognition algorithms*, Citeseer, 2011.

[7] S. Klum, H. Han, A. K. Jain, and B. Klare, "Sketch based face recognition: forensic vs. composite sketches," in *Proceedings of the 2013 international conference on biometrics (ICB)*, pp. 1–8, IEEE, Madrid, Spain, June 2013.

[8] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain, "Matching composite sketches to face photos: a component-based approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 191–204, 2012.

[9] B. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 639–646, 2010.

[10] B. Abaci and T. Akgul, "Matching caricatures to photographs," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 295–303, 2015.

[11] X. Tang and X. Wang, "Face sketch synthesis and recognition," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pp. 687–694, IEEE, Nice, France, October 2003.

[12] B. Klare and A. K. Jain, "Sketch-to-photo matching: a feature-based approach," *Biometric technology for human identification VII. International Society for Optics and Photonics*, Article ID 766702, 2010.

[13] P. C. Yuen and C. H. Man, "Human face image searching system using sketches," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 37, no. 4, pp. 493–504, 2007.

[14] S. Liu, D. Yi, Z. Lei, and S. Z. Li, "Heterogeneous face image matching using multi-scale features," in *Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB)*, pp. 79–84, IEEE, New Delhi, India, April 2012.

[15] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," in *Proceedings of the International Conference on Biometrics*, pp. 523–530, Springer, 2007.

[16] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proceedings of the CVPR 2011*, pp. 513–520, IEEE, Colorado Springs, CO, USA, June 2011.

[17] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: a survey," *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.

[18] S. Pramanik and D. Bhattacharjee, "Geometric feature based face-sketch recognition," in *Proceedings of the International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME-2012)*, pp. 409–415, IEEE, Salem, India, March 2012.

[19] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "Memetically optimized mcwld for matching sketches with digital face images," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1522–1535, 2012.

[20] H. K. Galoogahi and T. Sim, "Inter-modality face sketch recognition," in *Proceedings of the 2012 IEEE International Conference on Multimedia and Expo*, pp. 224–229, IEEE, Melbourne, Australia, July 2012.

[21] Z. Khan, Y. Hu, and A. Mian, "Facial self similarity for sketch to photo matching," in *Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA)*, pp. 1–7, IEEE, Fremantle, Australia, December 2012.

[22] H. Kiani Galoogahi and T. Sim, "Face photo retrieval by sketch example," in *Proceedings of the 20th ACM international conference on multimedia*, pp. 949–952, Nara, Japan, November 2012.

[23] H. S. Bhatt, S. Bharadwaj, R. Singh, and M. Vatsa, "On matching sketches with digital face images," in *Proceedings of the 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–7, IEEE, Washington, DC, USA, September 2010.

[24] H. K. Galoogahi and T. Sim, "Face sketch recognition by local radon binary pattern: LRBP," in *Proceedings of the 2012 19th IEEE International Conference on Image Processing*, pp. 1837–1840, IEEE, Orlando, FL, USA, October 2012.

[25] S. Pramanik and D. Bhattacharjee, "Geometric feature based face-sketch recognition," 2013, https://arxiv.org/abs/1312.1462.

[26] D. E. Goldberg, *Genetic algorithms*, Pearson Education India, Delhi, India, 2006.

[27] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories,"vol. 2, pp. 2169–2178, in *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2, pp. 2169–2178, IEEE, New York, NY, USA, June 2006.

[28] R. G. Uhl and N. da Vitoria Lobo, "A framework for recognizing a facial image from a police sketch," in *Proceedings of the CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–593, IEEE, San Francisco, CA, USA, June 1996.

[29] R.-L. Hsu and A. K. Jain, "Semantic face matching,"vol. 2, pp. 145–148, in *Proceedings of the IEEE International Conference on Multimedia and Expo*, vol. 2, pp. 145–148, IEEE, Lausanne, Switzerland, August 2002.

[30] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2008.

[31] D. Maturana, D. Mery, and A. Soto, "Face recognition with local binary patterns, spatial pyramid histograms and naive bayes nearest neighbor classification," in *Proceedings of the 2009 International Conference of the Chilean Computer Science Society*, pp. 125–132, IEEE, Santiago, Chile, November 2009.

[32] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[33] B. F. Klare, S. S. Bucak, A. K. Jain, and T. Akgul, "Towards automated caricature recognition," in *Proceedings of the 2012 5th IAPR International Conference on Biometrics (ICB)*, pp. 139–146, IEEE, New Delhi, India, April 2012.

[34] S. Ouyang, T. Hospedales, Y.-Z. Song, and X. Li, "Cross-modal face matching: beyond viewed sketches," in *Proceedings of the Asian Conference on Computer Vision*, pp. 210–225, Springer, Singapore, November 2014.

[35] Y. Yanwei Pang, X. Xuelong Li, Y. Yuan Yuan, D. Dacheng Tao, and J. Jing Pan, "Fast haar transform based feature extraction for face representation and recognition," *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 3, pp. 441–450, 2009.

[36] B. Klare and A. K. Jain, "On a taxonomy of facial features," in *Proceedings of the 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, pp. 1–8, IEEE, Washington, DC, USA, September 2010.

[37] F. R. Bach, "Consistency of the group lasso and multiple kernel learning," *Journal of Machine Learning Research*, vol. 9, no. 6, 2008.

[38] G. R. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *Journal of Machine Learning Research*, vol. 5, pp. 27–72, 2004.

[39] C.-C. Chang and C.-J. Lin, "Libsvm," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[40] A. C. Müller and S. Guido, *Introduction to machine learning with Python: A guide for data scientists*, O'Reilly Media, Inc., Sebastopol, CA, USA, 2016.