Migration Letters

Volume: 21, No: 3, pp. 1228-1239

ISSN: 1741-8984 (Print) ISSN: 1741-8992 (Online)
www.migrationletters.com

Centroid-Based Representation for Gender Classification Using Celebrity Cartoon Images

Prajna S[1], D S Guru[2], N Vinay Kumar[3]

Abstract

This work focuses on the exploitation of the notion of representing celebrity cartoon genders by clustering based on the cluster centroid. In this approach, features and samples are systematically reduced to provide a compact representation. For the extraction of deep features, the FaceNet architecture is used. The K-means algorithm is used to cluster celebrity cartoons based on their gender. The approach is carried forward to preserve the cluster representatives of each cluster for further classification and sample space reduction. We also explored suitable conventional classifiers and recommended the K-Nearest Neighbor (KNN) classifier, which is suitable for centroid-based gender classification. Further, the well-known subspace techniques such as Principal Component Analysis (PCA) and Fisher Linear Discriminant (FLD) have been adopted for dimensionality reduction. To demonstrate the effectiveness of the proposed model, we conducted extensive experiments on a dataset specific to celebrity cartoon images, namely IIIT-CFW. In contrast to existing methods, our model achieved a state-of-the-art result with an F-measure of 92.99%.

Keywords: *Deep features, clustering, centroid representation, celebrity cartoon gender classification.*

1. Introduction

Cartoon faces are widely recognized as an appealing form of art that is frequently used in many aspects of our daily lives. Cartoons are characterized by their non-realistic or semi-realistic depictions of real-world situations, often featuring amusing interpretations. Various fields, including entertainment, education, and advertising, use them extensively. In education, cartoons are a popular means of communicating complex ideas in a simple and engaging manner. Furthermore, they are widely used in advertising to capture people's attention and convey brand messages in a memorable and humorous manner. In the entertainment industry, cartoons are a cornerstone of animated shows and movies, providing viewers with hours of entertainment. There exists a significant modeling gap between these cartoons and the original images, as shown in Figure 1. The original celebrity image characters are associated with their celebrity cartoon characters, as depicted. Though recognizing these characters is a challenging task, recognizing gender in cartoons is even more difficult.

[1] Department of Studies in Computer Science, Manasagangorthi, University of Mysore, Mysuru 570006, India
[2] Department of Studies in Computer Science, Manasagangorthi, University of Mysore, Mysuru 570006, India
[3] Standard chartered Bank-GBS, Bangalore.

Figure.1: Illustration of original images with their associated celebrity cartoon characters to demonstrate the modelling gapbetweenoriginalandcartoonimages.

It is easier to determine gender in real images than in cartoons due to their intangible structural representation and appearance. However, gender classification in cartoons has been attracting the attention of many researchers in different fields, such as computer vision and machine learning, because of its potential applications. These applications include visual surveillance, intelligent human-computer interaction, social media analysis, demographic studies, and augmented reality. We focus on recognizing celebrity cartoon faces, which is challenging due to artistic variations, but even more difficult is recognizing the gender of celebrity cartoons. Accordingly, we use celebrity cartoons to address the problem of gender classification. However, identifying gender using these images is still more demanding, so we decided to take up the project with the intention of reducing the complexity of the search process. According to Figure 2, Aamir Khan and Narendra Modi exhibit high intra-class variations. There should be fewer intra-class variations and high inter-class variations in any classification model to be effective, but in our scenario, this is much more difficult.

(a) **(b)**

Figure 2: Exhibits intra-class variations a) samples belongs to class Aamir khan b) samples belongs to Narendra Modi.

Therefore, we attempted to categorize genders using recognized cartoon images.The classification of gender is a binary class problem that carries highly distinguished information concerning gender specifics. It aims to recognize the gender of a person based on the characteristics that differentiate between masculinity and femininity. Based on a review of the existing literature, we discovered that most of the studies involved the recognition of real images. However, the work has reported on the cartoon character recognition only. The current work aims to recognize celebrity cartoon images based on genders.

Meanwhile we discovered literature on identifying the gender of cartoon celebrities. In[3] used a multitask cascaded convolutional network (MTCNN) to detect faces and compared it to traditional methods. Contributions in this paper include using an inductive transfer learning approach that combines the feature learning capacity of the Inception v3 network and the feature recognition ability of support vector machines (SVMs) for face recognition. We also propose a hybrid convolutional neural network (HCNN) framework trained on a fusion of pixel values and 15 manually placed facial key points. The Inception+SVM model has been demonstrated to achieve a state-of-the-art (SOTAF1 score on the task of gender recognition of cartoon faces using a supportive unique dataset, namely the IIIT-CFW dataset.

With this backdrop, we attempted towards a cluster-based approach for celebrity cartoon face gender classification. Further, extensive experimentations are conducted to evaluate the performance of the proposed method by projecting the features onto Fisher space and eigen space. The experimentation results are evaluated using F1-Meaure as a metric.

The following contributionsof this paper are as follows:

- Considerable reduction is achieved in data representation by choosing the cluster representatives.
- Conduction of extensive experimentation and exploration of clustering-based approaches for gender in celebrity.
- Investigation of Suitable Classifier centroid-based gender classification.

The rest of the paper is structured as follows: In Section 2, a brief introduction of the proposed framework is presented. In Section 3, the details of experimental setup, dataset used are summarized. Results and comparative study have been presented in Section 4. Finally, Section 5 follows with conclusions.

2. Proposed Framework

In this section, a brief introduction to the proposed model (Figure 4) is given. The model comprises different stages, viz., feature extraction, clustering, feature selection, and classifier.

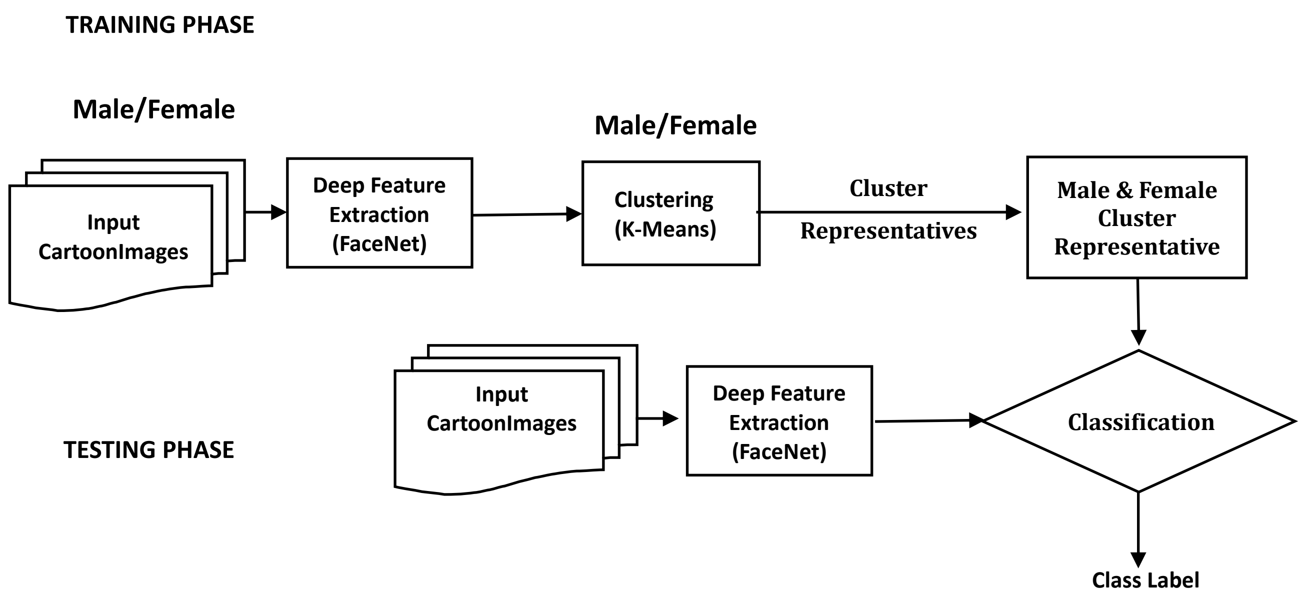


Figure 4: Architecture of Cluster-Based Approach

2.1Deep Feature Extraction

FaceNet architecture is a popular deep-learning model specifically designed for face recognition. This pre-trained model shows impressive performance on real-life face images. Thus, here in our work, we adopt the FaceNet architecture[23] for representing celebrity cartoon faces also. We use the pre-trained architecture, and train cartoon images of celebrities, generating distinct and deep embeddings with a triplet loss function. Which improves recognition accuracy. This motivated us to consider learning celebrity gender in cartoon images.

2.2 Cluster Centroid-based representation

K-Means clustering is a type of partitional clustering. Is adopted for clustering the celebrity cartoon faces of same gender. Hence instead of having a single larger higher space representation for a class, it is to have split multiple sub-higher spaces of same dimension. The general architecture of the cluster-based approach is shown in Figure 4. We recommend clustering the N_1 samples of the male class into K_1 clusters and the N_2 number of samples of the female class into K_2 clusters using K-means clustering. The K_1 and K_2 are fixed empirically by varying K under varying trails(see Fig.3.) to optimize the performance of the proposed gender classification method. Finally, the centroids of K_1 clusters of male class and centroids of K_2 clusters of female class is considered as centroid based representatives for classification.

2.3 Dimensionality Reduction

Feature dimension is reduced by adopting subspace methods called Principal Component Analysis (PCA) [30] and Fisher Linear Discriminant (FLD) [29] have been used to preserve the most dominating projection vectors. The contribution of eigenvalues and eigenvectors in the literature of face recognition has created a milestone; hence in this study, PCA and FLD have been adopted for dimensionality reduction, which indeed reduces the computation burden.

PCA: Is the Linear transformation technique. Let Y represents a linear transformation feature matrix mapping the function points from n-feature dimension to d-dimension, where d<<n, as follows:

$$Z_d = Y^T X_n \qquad (1)$$

In PCA, the projection Y_{opt} is chosen to maximize the determinant of the total scatter matrix of the projected samples, i.e.,

$$Y_{opt} = \arg\max \left| Y^T S_T Y \right| \qquad (2)$$

Where $\{Y_i | i = 1,2, \ldots, d\}$ is the set of n-dimensional projection vector corresponding to the d largest non-zero eigenvalues in eq.

LDA: A class specific approach is the Fisher's Linear Discriminant (FLD). This method selects the W in such a way that maximizes the ratio of the scatter between classes and the scatter within classes.

i.e.., W=arg max $\left| \frac{W^T S_B W}{W^T S_W W} \right|$ (3)

where S_B is between-class scatter matrix defined as $S_B = \sum_{i=1}^{c} N_i (x_i - \mu)(x_i - \mu)^T$and S_w is within-class scatter matrix defined as $S_W = \sum_{i=1}^{c} (x_i - \mu)(x_i - \mu)^T$ where c is the number of class $X_i$and N_i is the number of samples in class X_i. The $\{W_i | i = 1,2, \ldots \ldots, m\}$ is the set of generalization eigen vectors of S_B and S_w corresponding to the m largest generalized eigenvalues $\{\lambda_i | i = 1,2, \ldots \ldots m\}$.

2.4 Conventional Learning Algorithms

We have used five distinct supervised conventional learning models. In the literature, there are plenty of learning models used for face recognition, viz. Support Vector machines (SVM), K-nearest neighbor (K-NN), Decision tree (DT), Random Forest (RF), Gaussian Naive Bayes (GNB), (Kotsiantis, et., al 2006) and so on associated with this SVM with RBF kernel outperforms well in terms of classification metrics.

3 EXPERIMENTATION

3.1 Dataset

We have evaluated our model using the existing dataset [21]. The dataset consists of 8928 cartoon faces of 100 public celebrity figures, of which 6565 are male celebrity cartoon faces and 2073 are female celebrity cartoon faces, respectively. The samples that belong to the dataset are shown below in Figure 5.

Figure5: Pictorial representation of the Samples of IIIT-CFW dataset.

3.2 Experimental Setup and Results

During the stage of feature extraction, the FaceNet architecture, which is suggested for real faces, is adapted for extracting deep features. Using this pre-trained architecture, we extract 128-dimensional embeddings. Subsequently, class-specific clustering is carried out as explained above in Section 2.2 . Here, k_1 and k_2 values are fixed empirically by varying from 2 to 50 to fix up the suitable cluster value under a varying train-test ratio see in figure 3. For classification, as we explain in section 2.4 explored five distinct classifiers. During experimentation, the dimensionality reduction has been studied by adding the principal components sequentially and followed by FLD. Then, the cumulative match characteristic is calculated from each projection vector and its results are shown here.

Result Analysis of Suitable Classifiers

Initially all samples were considered for the study. We explored the performance of different classifiers from Figure 6 (i-iii) under varying training and testing ratios. It’s noticeable that SVM with RBF kernel outperforms the other classifiers under varying train test ratios. This is suitable for non-linear data as it transforms data points to fit the hyperplane between classes. Subsequently, experimentation results are evaluated with different trails (T = 10). The performance of the model is evaluated using F-Measure.

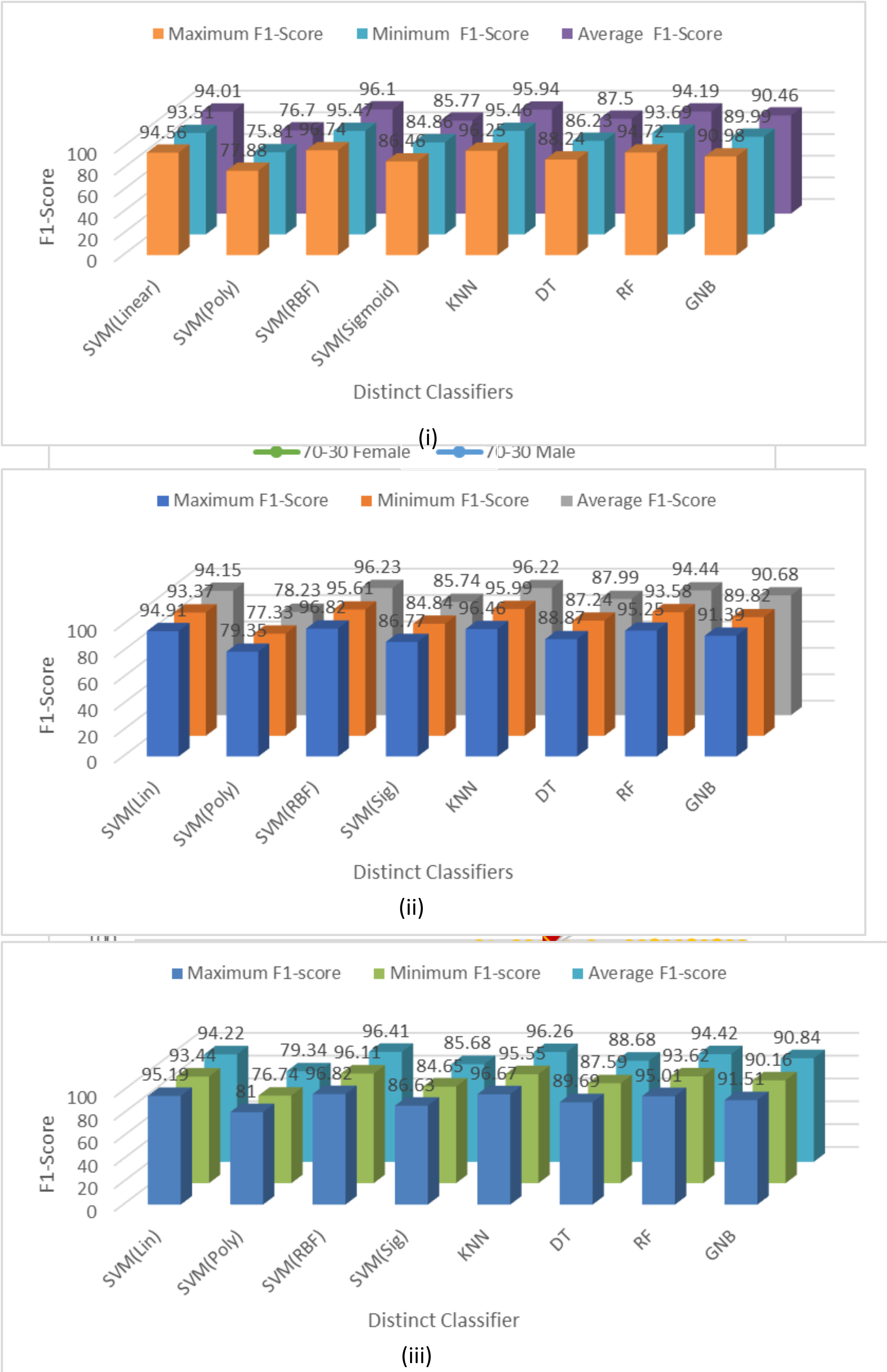
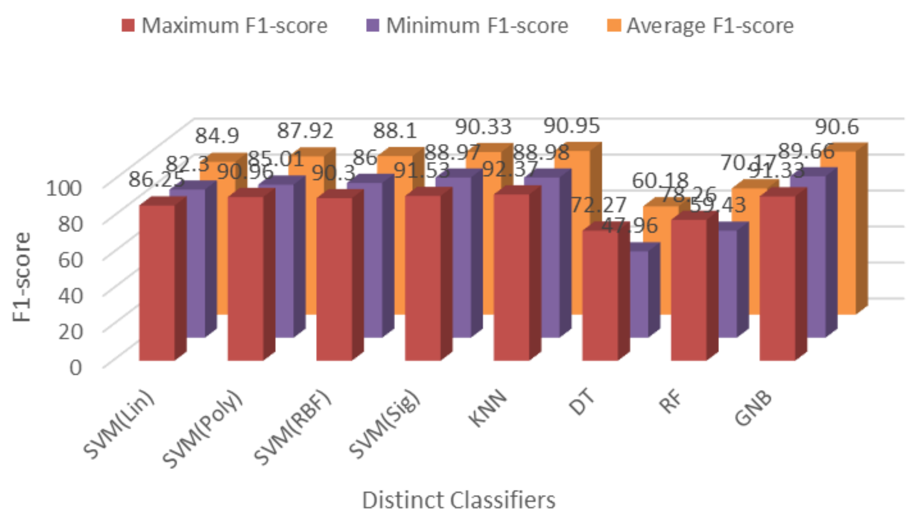


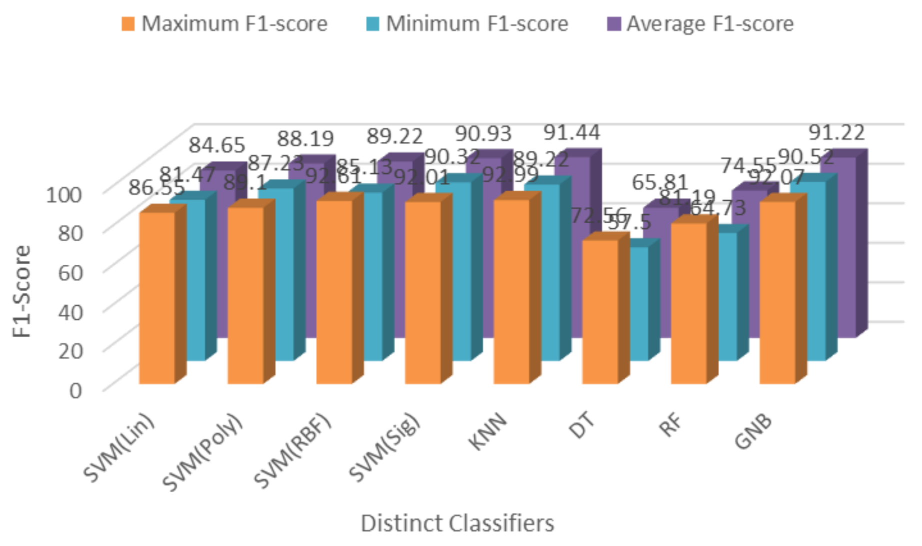
Figure 3: Consolidated K-values under varied train and test ratio (a) for 60:40, (b) for 70:30, and (c) for 80:20.

Figure 6:Performance of suitable classifiers under varying train-test ratio.(i) 60:40 train test ratio. (ii) 70:30 train-test ratio (iii) 80:20 train-test ratio

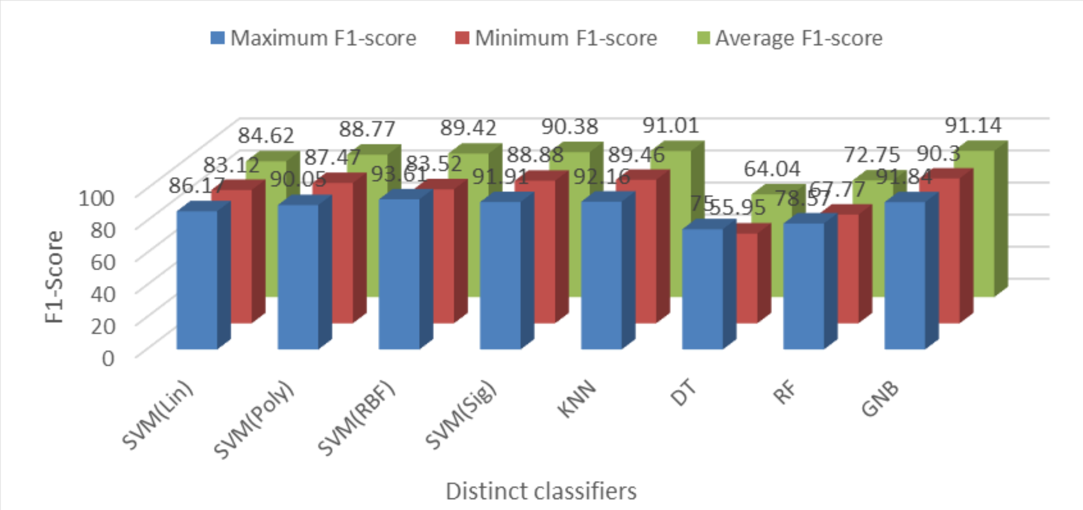
Similarly, we observed that the KNN classifier performs well after choosing clustering representatives. By choosing the cluster representatives, we achieve a considerable reduction in data representation is very effective for further classification. Empirically studied thatfor classification of the reduced sample space the K-Nearest Neighbor classifier outperforms than the other classifiers.



(i)



(ii)



(iii)

Figure 7: Performance results Obtain after clustering & Selection of class representative under varying train-test ratios.(i) 60:40 train test ratio. (ii) 70:30 train-test ratio (iii) 80:20 train-test ratio

It was observed that a maximum F1-score of 92.37 and an average F1-score of 90.95 were obtained for a 60:40 train-test ratio. On the other hand, a maximum F1-score of 92.99 and an average F1-score of 91.44 were obtained for the 70:30 train-test ratio. Similarly, for the 80:20 train-test ratio, a maximum F1-score of 92.16 and an average F1-score of 91.01 were observed. Based on these results, it can be concluded that the centroid-based representation displays remarkable performance.Further, we reduce the features in feature space by LDA and PCA transformers. It is observed that a compact representation is achieved by reducing both the sample and feature space indeed reduce the computational burden.Without cluster and with cluster transformation is studied and evident in Table 1.

4. Comparative Analyses

To evaluate further the recommended model, a comparative study is performed in following two distinct folds:

i. Across the proposed model

ii. Against the contemporary models

4.1 Across the proposed model

a. To emphasize our model's performance, we initially attempted to check the suitability of the classifier by comparing it against other conventional classifiers. The performance of the classifiers is demonstrated using Figure 6(i-iii). It is evident through extensive experimentation. The support vector machine (SVM) with RBF kernel outperforms with a maximum, minimum, and average F1-Score.

b. After applying k-means, and empirically fixing up the k-value, we explore distinct classifiers, and it is seen from Figure 7(i-iii) that K-NN is outperformingwith a maximum, minimum, and average F1-score of 92.16%, 89.46%, and 91.01%, respectively.

c. Further, we have compared our model by feature transformation techniques such as Linear Discriminant Analysis (LDA) and Principal component analysis (PCA) performance is analyzed. It is evident in Table 1. The result without clustering is highly dominating than with clustering but the main objective is that we achieve both numerosity and feature reduction with highly contributing samples. The results are promising.

Table 1. Consolidated results obtained the Average and Maximum F1-score from distinct methods.

Split	Approach	Transformation	Classifier	Average F1-score	Number of Features selected	Classifier	Maximum F1-score	Number of Features selected
60:40	Without Clustering	LDA	SVM (RBF Kernel)	87.25	59	SVM (RBF Kernel)	86.66	123
		PCA	KNN	95.69	20	KNN	96.47	75
	With Clustering	LDA	SVM (Polynomial)	64.85	38	SVM (RBF Kernel)	69.70	90
		PCA	SVM (RBF Kernel)	83.13	29	SVM (RBF Kernel)	92.17	43
70:30	Without Clustering	LDA	SVM (RBF Kernel)	89.4	54	SVM (RBF Kernel)	91.31	124
		PCA	KNN	89.76	16	KNN	95.98	85
	With Clustering	LDA	SVM (Polynomial)	59.62	27	SVM (Polynomial)	70.33	81
		PCA	SVM (RBF Kernel)	84.96	30	SVM (RBF Kernel)	86.44	80
80:20	Without Clustering	LDA	SVM (RBF Kernel)	89.96	58	SVM (RBF Kernel)	91.61	119
		PCA	KNN	92.46	20	KNN	96.41	89
	With Clustering	LDA	SVM (Polynomial)	60.70	51	SVM (Polynomial)	69.40	70
		PCA	SVM (RBF Kernel)	83.46	29	SVM (Polynomial)	85.23	90

4.2 Against the contemporary models

We compared our model with existing other models to emphasize the superiority of our model. It is also evident from Table 2. Centroid based approach achieves F1-score of 92.99 than the existing F1-score.

Table2: Performance comparison of proposed model with respect to other existing models for gender classification using celebrity cartoon images.

ProposedWork							
Model	Precision		Recall		F1-score		BalancingData
	MIN Avg	MAX Avg	MIN Avg	MAX Avg	MIN Avg	MAX Avg	
Centroid-BasedApproach	90.55	93.35	88.56	92.86	88.98	92.99	NO
ExistingWork							
HCNN (Jhaet.al.,2018)	90.40		82.70		86.40		YES
INCEPTION V3+SVM (Jha et.al.,2018) (2048 Features)	92.70		89.40		91		YES

5. Conclusion

In this paper,we made a successful attempt towards classification of gender based on the cluster centroids. We have adopted clustering method to group similar cartoon face images of the same gender.Further, for each cluster of cartoon face images, consider a centroid based representation. A suitable conventional classifier is studied and adapted for classification. The added advantage of theproposed method is that it has effective and efficient representation of gender cartoon faces.An extensive experimentation has been conducted to demonstrate the effectiveness of the proposed approach and a comparative study reveals that the proposed approaches superiors other existing contemporary models.

References

1. Li, L., Mu, X., Li, S., & Peng, H,:A review of face recognition technology. IEEE access, 8, 139110-139120, (2020).
2. Arai, M.,: Feature extraction methods for cartoon character recognition. In 2012 5th International Congress on Image and Signal Processing (pp. 445-448). IEEE, (2012).
3. SauravJhaandNikhilAgarwalandSuneetaAgarwal,.: TowardsImprovedCartoonFaceDetectionandRecognitionSystems,CoRR,vol. abs/1804.01753, (2018).
4. Jing Huo, Wenbin Li, Yinghuan Shi, Yang Gao, and Hujun Yin.,.:WebCaricature: a benchmark forcaricaturefacerecognition,arXivpreprintarXiv:1703.03230, (2017).
5. Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., & Aizawa, K.,.:Manga109 dataset and creation of metadata. In Proceedings of the 1st international workshop on comics analysis, processing and understanding, (pp. 1-5),(2016).
6. Branwen, G., &Gokaslan, A.,.:Danbooru2019: A large-scale crowdsourced and tagged anime illustration dataset. Danbooru2017, (2019).
7. Abaci, B., & Akgul, T.,:Matching caricatures to photographs. Signal, Image and Video Processing, 9, 295-303, (2015).

8. Chen, H., Zheng, N. N., Liang, L., Li, Y., Xu, Y. Q., & Shum, H. Y.:PicToon: a personalized image-based cartoon system. In Proceedings of the tenth ACM international conference on Multimedia (pp. 171-178), (2002).

9. Deng, J., Guo, J., Xue, N., & Zafeiriou, S.,:.Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 4690-4699), (2019).

10. Elad, M., Starck, J. L., Querre, P., & Donoho, D. L.,: Simultaneous cartoon and texture image inpainting using morphological component analysis (MCA). Applied and computational harmonic analysis, 19(3), 340-358., (2005)

11. Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D.,:. The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 38(1), 149-161, (2007).

12. Guo, Y., Zhang, L., Hu, Y., He, X., & Gao, J.:. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14 (pp. 87-102). Springer International Publishing, (2016).

13. He, K., Zhang, X., Ren, S., & Sun, J,.:. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778), (2016).

14. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q, .:. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4700-4708), (2017).

15. Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E.,:. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Workshop on faces in'Real-Life'Images: detection, alignment, and recognition,(2008).

16. Jing Huo, Wenbin Li, Yinghuan Shi, Yang Gao, and HujunYin,.:WebCaricature: a benchmark forcaricature facerecognition.arXivpreprintarXiv:1703.03230(2017).

17. Kemelmacher-Shlizerman, I., Seitz, S. M., Miller, D., & Brossard, E,.:Themegaface benchmark: 1 million faces for recognition at scale. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4873-4882), (2016).

18. Kendall, A., Gal, Y., & Cipolla, R,.: Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7482-7491), (2018).

19. Klare, B. F., Bucak, S. S., Jain, A. K., & Akgul, T, .: Towards automated caricature recognition. In 2012 5th IAPR international conference on biometrics (ICB) (pp. 139-146), IEEE, (2012).

20. AshutoshMishra,ShyamNandanRai,AnandMishra,andCVJawahar.: IIITCFW:abenchmarkdatabaseofcartoonfaces inthewild.InEuropeanConference onComputerVision.Springer,35–47, (2016).

21. A.S.Dhavalikar andR. K. Kulkarni, "Facedetectionandfacialexpressionrecognition system," 2014InternationalConferenceonElectronicsandCommunicationSystems(ICECS),2014,pp.1-7,doi:10.1109/ECS.2014.6892834.

22. C.DingandD.Tao,"RobustFaceRecognitionviaMultimodalDeepFaceRepresentation,"in IEEETransactionsonMultimedia,vol.17,no.11,pp.2049-2058,doi:10.1109/TMM.2015.2477042, (2015).

23. F.Schroff,D.KalenichenkoandJ.Philbin,.: FaceNet:Aunifiedembeddingforfacerecognitionandclustering,pp.815-823, IEEEConference onComputerVisionand PatternRecognitionCVPR,(2015).

24. Anil K. Jain and Richard C. Dubes. Algorithms for Clustering Data. Prentice-Hall, Inc., Upper Saddle River,NJ,USA,(1988).

25. K. P. Sinaga and M. -S. Yang, "Unsupervised K-Means Clustering Algorithm," in IEEE Access, vol. 8, pp.80716-80727,2020,doi:10.1109/ACCESS.2020.2988796.

26. K.Taunk,S.De,S.VermaandA.Swetapadma,"ABriefReviewofNearestNeighborAlgorithmfor Learning and Classification," 2019 International Conference on Intelligent Computing and Control Systems (ICCS),Madurai,India,2019,pp.1255-1260,doi:10.1109/ICCS45141.2019.9065747.

27. AshishChittoraandOmMishra,"FaceRecognitionUsingRBFKernelBasedSupportVector Machine,"InternationalJournal ofFutureComputer andCommunicationvol.1,no.3,pp.280-283,2012.

28. Guru, DevanurS. and N. Vinay Kumar. “Symbolic Representation and Classification of Logos.” ArXiv abs/1612.08796 (2016): n. page.

29. Mika, S., Ratsch, G., Weston, J., Scholkopf, B., Mullers, K.R.: Fisher discriminant analysis with kernels. In: Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop (cat. no. 98th8468), pp. 41–48. IEEE, August 1999.

30. Turk, M.A., Pentland, A.P.: Face recognition using Eigen faces. In: Proceedings CVPR’91, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE (1991).