Face Image-Based Gender Classification of Children

R. Sumithra[1](✉), D. S. Guru[1], and Manjunath Aradhya[2]

[1] Department of Studies in Computer Science, University of Mysore, Manasagangotri, Mysuru 570006, Karnataka, India
sumithraram55@gmail.com, dsg@compsci.uni-mysore.ac.in
[2] Department of Computer Application, JSS Science and Technology University, Mysuru, Karnataka, India
aradhya@sjce.ac.in

Abstract. In this work, we have addressed the problem of Gender Classification in children which is a challenging area in the face recognition system. We have adopted two alternative approaches by varying feature extraction techniques: conventional and convolutional techniques. In the conventional technique, the local and texture features such as HOG, M-LBP, LGP, and Histogram are taken. Subsequently, feature fusion is also performed. In the convolutional technique, the pre-trained deep neural network models such as AlexNet, GoogLeNet, VGG-19, and ResNet-101 are utilized for feature extraction. For the classification task, standard learning models and their ensemble are tried out. The proposed model has experimented on our longitudinal data of toddlers (four years) to preteen age (fourteen years) consisting of 56000 face images from 450 children. During the experimentation, the performance of gender classification with age information and without age information is evaluated. Finally, group-based gender classification is also estimated for further analysis of the model. From this extensive experiment, we have observed that gender classification of young children concerning age is quite challenging than children with groupage information. The results from deep features achieved 99.1% of F-Measure for group-based gender classification of children.

Keywords: Children face data · Deep convolution neural network · Gender classification · Group age analysis

1 Introduction

Face recognition is an interesting area, and much work has been attempting to help face images in the wild for pose variation, facial reconstruction identification, illumination variation, etc., (Jain et al., 2016). The fundamental characteristic of identity recognition is the human face. The face photos are useful not only for the identification of individuals but also for exploring other characteristics such as gender, age, race, a person's emotional state, etc. A face biometric that allows it a suitable modality due to its individuality, universality, acceptance, and ease of collectability. Various problems on face recognition

D. S. Guru et al. (Eds.): ICCR 2021, CCIS 1697, pp. 213–228, 2022.
https://doi.org/10.1007/978-3-031-22405-8_17

have been studied for the last few decades and several works have been reported in the literature. The major contributions are found based on eigenfaces (Belhumeur et al., 1997; Turk and Pentland, 1991), fisher faces (Craw et al., 1999; Wright et al., 2008), and neural network-based (Lawrence et al., 1997; Wang et al., 2018). In decades ago, the face recognition system was analyzed and evaluated by Bledsoe (1966). Sparse Representation Coding (SRC) (Wright et al., 2008) and deep learning models (Sun et al., 2014) are some of the most notable advances in the area of face recognition. Face recognition using subspace techniques has been effectively studied by Rao and Noushath (2010).

Along with face recognition, gender classification is an equally important area related to the aging problem. Hidden-Markov Model with support-vectors was used for representing face patch distributions (Zhuang et al., 2008). Guo et al. (2009) were studied age estimation which is a combination of Biologically-Inspired Features (BIF) and multiple manifold-learning techniques. Gabor and local binary patterns (LBP) features, as well as a hierarchical age classifier composed of Support Vector Machines (SVM) to classify the input image into an age-class, followed by a support vector regression, were used to estimate an accurate age of a person (Choi et al., 2011). A detailed survey of gender classification methods can be found in Makinen and Raisamo (2008). One of the early methods for gender classification (Golomb et al., 1990) used a trained neural network on a small set of near-frontal face images. In (O'toole et al., 1997), the combined 3D structure of the head (obtained using a laser scanner) and image intensities were used for classifying gender. SVM classifiers were used by Moghaddam and Yang (2002), applied directly to image intensities. Recently, (Ullah et al., 2012) used Weber's local texture descriptor for gender recognition, demonstrating near-perfect performance on the FERET benchmark dataset. Finally, viewpoint invariant age and gender classification was presented by Toews and Arbel (2008). Deep Convolutional Neural Network (D-CNN) for gender recognition based on each facial component has been studied on adults (Lee et al., 2019). Gender classification on speech data using Gaussian mixture models (GMM) and Kullback–Leibler divergence for learning model. Yücesoy (2020) achieved a classification accuracy of 92.3%. The maximum work was found on UIUC-IFP-Y, FERET, FG-NET, and MORPH datasets on adults. However, the FG-NET dataset consists of both adult's and children's face images. But, in the literature, there was no much work has been accomplished on the gender classification of children. The accuracy of the gender classification is largely depending on age. Hence, the literature which has been covered in this work is very limited.

Due to the lack of children's face data and efficient models, the classification of children's gender is an open problem in the face recognition community. The application of the existing human-computer system such as passive surveillance and collection of demographics, age prediction, and gender classification in adults has grown into a comprehensive field of research. Child age prediction and gender classification is also a growing area in the applications of automatic school attendance marking systems, proper vaccination tracking, auto-renewal of Government IDs, etc. During this research, we have created a longitudinal face image dataset of children to address numerous problems, such as identifying the stable age of a person, aging recognition, accurate age prediction, and gender classification of children. Our dataset has face images of very young children from age four to fourteen years. Hence, the accurate age prediction and gender classification in very young children is challenging. This work focuses on gender classification problems. To show, how these young children's face image patterns are distributed for the aforementioned problems, suitable and efficient approaches have been well addressed. We have adopted two alternative approaches by varying feature extraction techniques: conventional and convolutional techniques. Trained a multi-classification algorithm using the ensemble technique. During experimentation, the performance of gender classification with age information and without age information is performed. Finally, group-based gender classification is also computed for further analysis. Along with dataset creation, this work has a significant contribution to children's gender and group-based gender classification. The comparison between conventional and convolutional methods also enhances the novelty of this work.

The remaining part of this paper is organized as follows: Sect. 2 provides the proposed framework for gender classification model with pre-processing, feature extraction techniques, and classification algorithms. Section 3, consisting of the dataset creation, and Sect. 4 provides experimental results and analysis. Section 5 follows with a conclusion.

2 The Proposed Method for Children Gender Classification

The proposed model is designed using four stages viz., pre-processing, feature extraction, feature fusion, and learning models, as depicted in Figs. 1 and 2. Initially, the input images have been pre-processed by face alignment. Further, features have been described from the facial components of the images to represent the facial parameters and followed by feature fusion. The different learning models have been utilized for classification, and the obtained results are finally fused using an ensemble technique, for further analysis of our data.

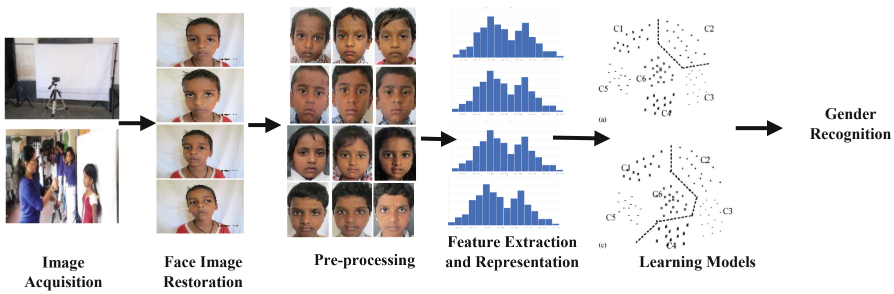

Fig. 1. Workflow of the proposed Gender Classification model.

2.1 Image Pre-processing and Normalization

In pre-processing, the captured children's face images have gone with image localization to maintain consistency across the images. Hence, the face has been detected using the Viola-Jones detection technique (Viola and Jones, 2001) and normalized by using histogram equalization, which is depicted in Fig. 3.

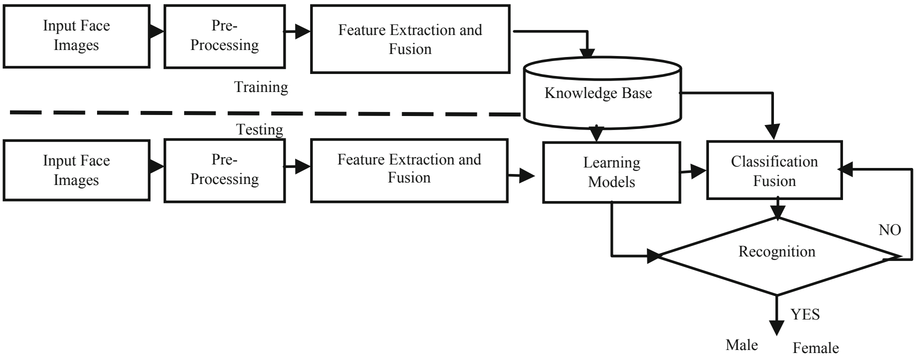

Fig. 2. Flowchart of our proposed Gender Classification model.

(a) (b)

Fig. 3. The Illustration of pre-processing (a) captured face images; (b) pre-processed images.

2.2 Feature Representation: Conventional and Convolutional

The features which are used to characterize the face images of children are described. The patterns of children's face images are distributed smoothly. In the literature, local features are widely demanded as their efficient and effective discriminant in representing facial features. Therefore, we have adopted local, and texture features such as Multi-scale Local Binary Pattern (M-LBP) (Ahonen et al., 2006), Histogram of Orientation Gradient (HOG) (Dalal and Triggs, 2005), Local Gabor Pattern (LGP) (Zhang et al., 2005) and histogram are used to preserve the face features. Along with the local and handcrafted features, pre-trained weights in deep learning-based features have been extracted from the processed images for different feature comparisons. To show the state-of-art technique, Deep features from the pre-trained networks such as AlexNet (Krizhevsky et al., 2012), GoogLeNet (Szegedy et al., 2014), VGG-19 Layers (Simonyan and Zisserman, 2014), and ResNet-101 layers (He et al., 2016) have also been adopted.

Multi-scale Local Binary Pattern (M-LBP): LBP operator is one of the best performing texture descriptors (Ahonen et al., 2006). In this method, the operator works by thresholding a 3×3 neighborhood with the value of the center pixel, thus forming a local binary pattern, which is constructed as a binary number. The operator is extended to facilitate rotation invariant analysis of facial textures at multiple scales such as 3×3, 5×5, 7×7, and 9×9 (Ojala et al., 2002). From the face image of size 500×500, we have acquired with 59-d feature vector; therefore, a total of 59×5 (multi-scale) = 295-d feature vectors are extracted from the MLBP feature.

Histogram of Orientation Gradient (HOG): The HOG feature is extracted by splitting the face image into small cells (Dalal and Triggs, 2005). The cells in each detected window are 16 × 16 pixels in size, and each 8 × 8 cell group is built in the sliding mode of a block. So that the blocks overlap with 50% from each cell. Every cell has a 9-bin histogram, each containing a concatenated vector of all cells. A vector of 36 D function, which is uniform to unit length, is represented for each block.

Local Gabor Pattern (LGP): Given a child's face image, its features are first extracted by converting them with several Gabor filters on different scales and directions. LBP is then used to encode the Gabor functionality's micropatterns. The Local Gabor Pattern (LGP) operator is the mixture of Gabor and LBP operators. The description of the LGP operator is:

$$\mathrm{LGP}_{\vartheta,\mu}(\mathrm{x_c}, \mathrm{y_c}) = \sum_{p=0}^{7} S(\mathrm{O}_{\psi\vartheta,\mathrm{m}}(\mathrm{XP}, \mathrm{y_p}) - \mathrm{O}_{\vartheta,\mathrm{m}}(\mathrm{x_c}, \mathrm{y_c}))2^{\mathrm{p}} \ldots \tag{1}$$

where $\psi\vartheta$, μ is the Gabor response, ϑ and μ are the scale and orientation variables respectively, as shown in Eq. 1. The histogram has been used to compile instances of various LGP trends images. To prevent the loss of spatial information by the histogram, the LGP image is divided into multi-region that do not overlap, and each sub-region extract histogram. Every histogram is combined to reflect the given face image in a single histogram. Considering Eq. 2, any LGP image is partitioned into different regions, H υ, μ, r representing (υ, μ, r) the histogram. Then, the final face representation by LGP can be denoted as:

$$\mathrm{H} = \left(\mathrm{H}_{0,0,0}, \ldots, \mathrm{H}_{0,0,\mathrm{n}-1}, \mathrm{H}_{0,1,0}, \ldots, \mathrm{H}_{0,1,\mathrm{n}-1}, \ldots, \mathrm{H}_{4,7,\mathrm{n}-1}\right) \ldots \tag{2}$$

AlexNet (Krizhevsky et al., 2012): To reduce the computational complexity of AlexNet pre-trained deep neural network architecture, the input face images are down-sampled from 500 × 500 to 227 × 227 in terms of spatial resolution. The proposed system employs five convolutional layers, three pooling layers, and rectified linear unit layers (ReLU). The first -convolution layer uses 96 kernels of relatively large size 11 × 11 × 3, while the second -convolutional layer uses 256 kernels of size 5 × 5. 384 kernels of size 3 × 3 are utilized in the third, fourth, and fifth levels. Each convolutional layer generates a feature map. The architecture has eight layered designs with a total of 4096 nodes, each node is treated as a descriptor.

GoogLeNet (Szegedy et al., 2014): Also known as the Inception module, plays an important role in GoogLeNet Architecture. To feed the child face images to GoogLeNet architecture, images are down-sampled from 500 × 500 to 224 × 224. Initially in the Inception of Architecture is restricted to the filter sizes 1x1, 3 × 3, and 5 × 5. A 3 × 3 max pooling is also added to the inception architecture. The network is 22 layers deep. The initial layers are simple convolutional layers. This network has 57 layers of inception module, among which 56 are convolutional layers and one fully-connected layer.

VGG-19 Layers (Simonyan and Zisserman, 2014): The input child face images of size 224 × 224 are passed through a stack of convolutionary layers in this architecture, in which filters with very small receptive fields are used: 3 × 3. A stack of convolutionary layers is followed by three fully connected (FC) layers: the first two have 4096 channels each, the third one performs 1000 ways, the configuration of the fully connected layers is the same across all networks. All hidden layers are equipped with rectification non-linearity. This network has 19 layers with weights, among which 16 are convolutional, and the remaining 3 have a fully connected layer.

ResNet-101 Layers (He et al., 2016): The convolutional layer have 3 × 3 filters and follow two simple design rules: (i) the layers have the same output feature size and (ii) the number of filters is doubled if the feature map size is halved to preserve the time complexity per layer. The network ends with a global average pooling, a 10-way fully connected layer, and softmax. This architecture has 105 layers with weights, among which 104 are convolutional and one fully connected layer.

The fusion of multi-features is performed to increase the rate of recognition and fused using Min-Max normalization. The fusion is taken for conventional features.

2.3 Baseline Learning Models

In literature, there are many learning models used for face recognition systems; in this study, we have evaluated the dataset's quality by conducting several baseline learning models. Different classification methods viz., K-Nearest Neighbor (K-NN) (Keller et al., 1985), RBF- Support Vector Machine (SVM) (Keerthi et al., 2001), Linear Discriminant Analysis (LDA) (Balakrishnama and Ganapathiraju, 1998) and Decision tree (Friedl and Brodley, 1997), since, we have utilized standard and available learning models the detailed description is not provided here.

Ensemble: An ensemble of classifiers is a group of classifiers whose individual judgments are merged in some way to classify new samples (usually through weighted or unweighted voting) (Quinlan, 1996). The approaches for creating good ensembles of classifiers have been one of the most active research areas in supervised learning. The simplest kind of majority voting is hard voting. Here, we have used the majority (plurality) voting of each classifier to forecast the class label Y:

$$\mathrm{Y} = \mathrm{Mode}\{\mathrm{SVM(x)},\ \mathrm{k\text{-}NN(x)},\ \mathrm{LDA(x)},\ \mathrm{Decision\ Tree(x)},\ \mathrm{Nave\ Bayes(x)}\}\ldots \tag{3}$$

where ‘x’ is the class labels.

3 Dataset Creation

The proposed model has been validated on a reasonably sized dataset created during June 2017–December 2019. In this section, we have presented detail on the dataset's creation and the number of images used for experimentation. We created our dataset of the longitudinal face images of young children of age 4 to 14 years. Our longitudinal face data collection was conducted in five different Government schools in and around Mysore, India. We have captured toddles to preteen age's longitudinal face images for 10 different sessions over 30 months (July 2017 to December 2019), every three months intervals. Data collection in each session captured 10–12 face images of every child subsequently over approximately 2 min. Approximately, 56000 images from 450 children were taken for dataset creation. Face images were captured in the school premises with a suitable setup made. To maintain a degree of consistency throughout the dataset, the same physical setup and location with a semi-controlled environment were used in each session. However, the equipment had to reassemble for each session; therefore, there was a variation from session to session. Images were captured in profile view with the varying pose, scaling, and angle, as illustrated in Fig. 4.

Fig. 4. Illustration of children dataset for gender classification.

Our gender classification model has been performed with different sets of experimentation, viz., gender classification of children from 4 to 14 years old (without age information)as shown in Table 1, gender classification of children (with age information)as shown in Table 2, and gender classification of children from different age group as shown in Table 3.

Table 1. Description of our datasets used for Gender Classification (GC) without age.

Gender Classification without age	
Male (children)	Female (children)
27,830 (220)	28,175 (228)

Along with the age-based gender classification, group-based gender classification is also performed. Three groups of ages are computed as Group-1 have an age interval of [4–5.9] years old children called pre-school children, Group-2 have an age interval of [6–9.9] years old children called primary education children and Group-3 have an age

Table 2. The detailed description of our datasets used for Gender Classification (GC) with age.

	Male (children)	Female (children)	Male (children)	Female (children)	Male (children)	Female (children)
Gender Classification with age	**4 years**		**5 years**		**6 years**	
	2167 (19)	2806 (28)	1836 (15)	2834 (23)	3847(36)	3878 (33)
	7 years		**8 year**		**9 years**	
	3745(30)	4136 (36)	3842 (33)	3923 (34)	4591 (36)	4139 (32)
	10 years		**11 years**		**12 years**	
	5540 (45)	3424 (25)	1138 (10)	2000 (15)	1127 (8)	1175 (11)

interval of [10–14.9] years old children called lower secondary education children. Since we have a longitudinal dataset, a child may participate in Group-1 may also participate in Group-2. The age of each child is computed by the difference in date of birth and date of the image captured.

Table 3. Description of our datasets used for Group-based Gender Classification (GC).

Number of images	Group Gender Classification	
	Male (children)	Female (children)
Group-1	4003 (95)	5640 (79)
Group-2	16025 (111)	15936 (66)
Group-3	7805 (53)	6599(49)

4 Experimentation and Results

Gender classification is a two-class problem: Male class and Female class. Our interest is to bring out the effectiveness of our novel dataset through rigorous experimentation on various forms. Two different forms of feature extraction modules are used, followed by baseline learning models. The K-NN classifier is used with a K value of 5 and Euclidean distance metrics for uniform weight adjustment of the model. RBF-SVM with a gamma value of 0.1 is used to learn the model for the feature matrix. Gini impurity for information gain in random of 2 minimum split of the tree is taken for decision tree classifier.

In deep learning-based classification, pre-trained neural network weights such as AlexNet, GoogLeNet, ResNet-101, and VGG-19 for gender classification problems have been utilized. During experimentation, resized the images as per the standardization of the deep pre-trained models. We have trained our network using the training set up by having the maximum epochs of 20 with a minimum batch size of 100, optimizer as

Stochastic Gradient Descent with Moment (SGDM), and learning rate of 0.0001 with 50% of training images, which are empirically analyzed. The feature dimension for AlexNet is 4096 with an input image size of 227 × 227, GoogLeNet is 1000 with an input image size of 224 × 224, ResNet-101 is 1000 with an input image size of 224 × 224, and VGG-19 is 4096 with an input image size of 224 × 224 with soft maxing in the last layer. To evaluate the model, we have conducted three sets of experiments. In the first experiment, we used 50% of the samples from each class to create class representative vectors (training phase) and the remaining 50% of samples for testing. The number of training and testing samples in the 60:40 and 70:30 ratios are the remaining set of experiments. For each collection of training and testing, we have randomly performed experimentation of 10 trails. The model's performance is measured using an average of precision, recall, F-Measure, and accuracy of the confusion matrix from the ten predicted results. Only the F-Measure for the 60:40 train and test split of the model is shown in each approach.

The results on gender classification for various approaches are presented from Figs. 5, 6, 7, 8, 9 and 10. Figure 5, shows the F-Measure of multi-classifiers for local features in individual ages of children from 4 to 12 years old. Similarly, Fig. 6 shows an F-Measure from deep features. Figure 7, shows the F-Measure of multi-classifiers for local features without age information of children from 4 to 12 years old. Similarly, Fig. 8 shows an F-Measure for deep features. Figure 9 shows an F-Measure from local features for three age groups of children; again, Fig. 10 shows an F-Measure from deep features.

The proposed models achieve promising results for the classification of gender in young children. During the experimentation, the multiple face images of the same children with a different period are taken. Hence, it impacts our obtained results. The results from the conventional method perform better for K-NN and LDA classifiers than SVM, this is due to the large sample size with normalized feature descriptors. The performance of SVM is very poor for all the experiments. Hence, the non-learning learning model is not fit for this problem. But HOG features give the best results irrespective of classifiers. The F-Measure for children of age 11 and 12 gives 90% off almost all the methods except SVM and LDA classifiers. However, an age after 11 years may be feasible to recognize the children's gender.

For both the experiments, the proposed method achieves the best classification rate, in terms of F-Measure. The deep features archives almost 100% of F-Measure for K-NN and ensemble classifiers. From the obtained results it is observed that deep features are more efficient than local or handcrafted features for gender classification of children WRT to different ages. All the deep features behavethe same for all classifier, hence the obtained results largely depends on the learning models.

Deep learning is the data hunger method since we have collected a large number of images, hence it performed well for our experimentation.Ensemble classifier for both local and deep features is performed well for gender classification without age information, the F-Measure reached nearly 100% for 70% of training data, refer Figs. 7 and 8. Referring to Figs. 9 and 10, it is difficult to analyze for the group-based gender classification problem, the two approaches provided almost the same F-Measure.

In group-based gender recognition, the F-Measure for children for Group 2 and Group 3 achieves best for both conventional and convolutional features. Hence, group

Fig. 5. F-Measure obtained from local features for gender classification with age

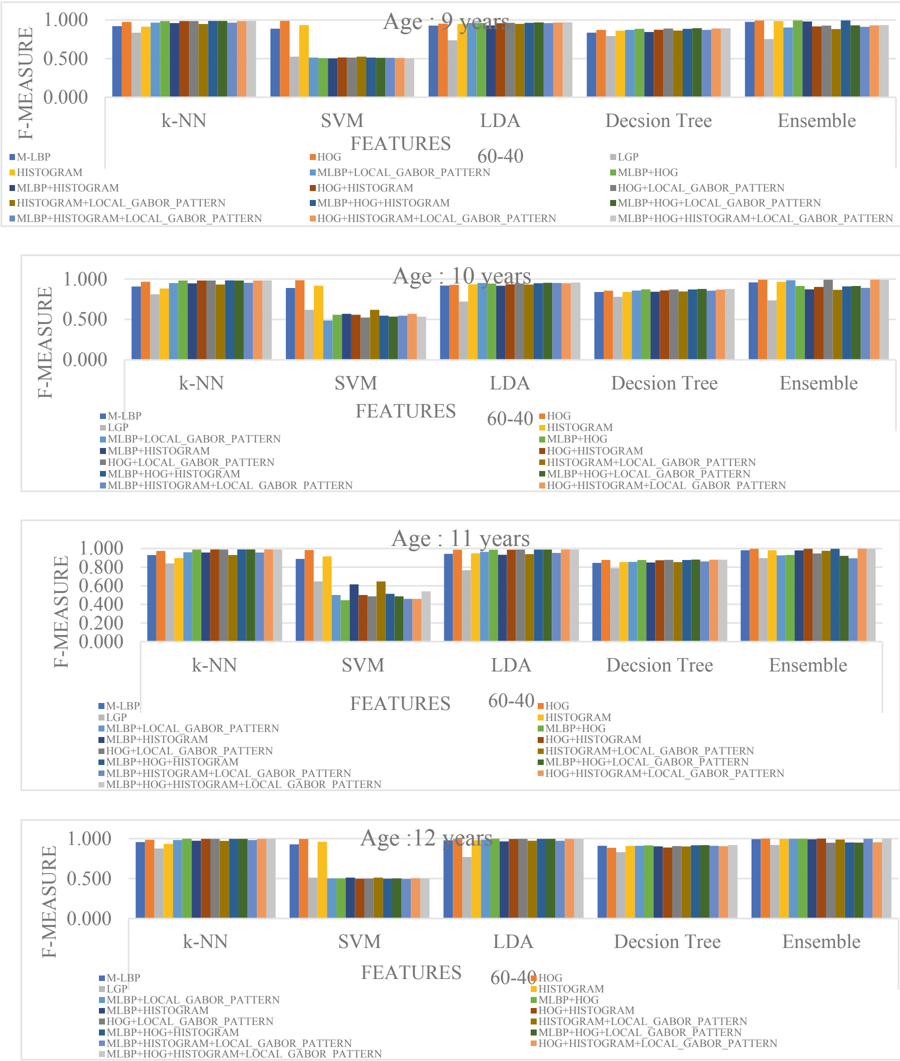

Fig. 5. (*continued*)

gender recognition of children after the age of 6 years may be easier. Among the three sets of problems, gender classification without age performed perform best F-Measure compared with the other two. From this extensive experiment, we have observed that gender classification of young children concerning age is quite challenging than children with groupage information. The results from deep features achieved 99.1% of the F-measure for group-based children gender classification, which is appreciated.

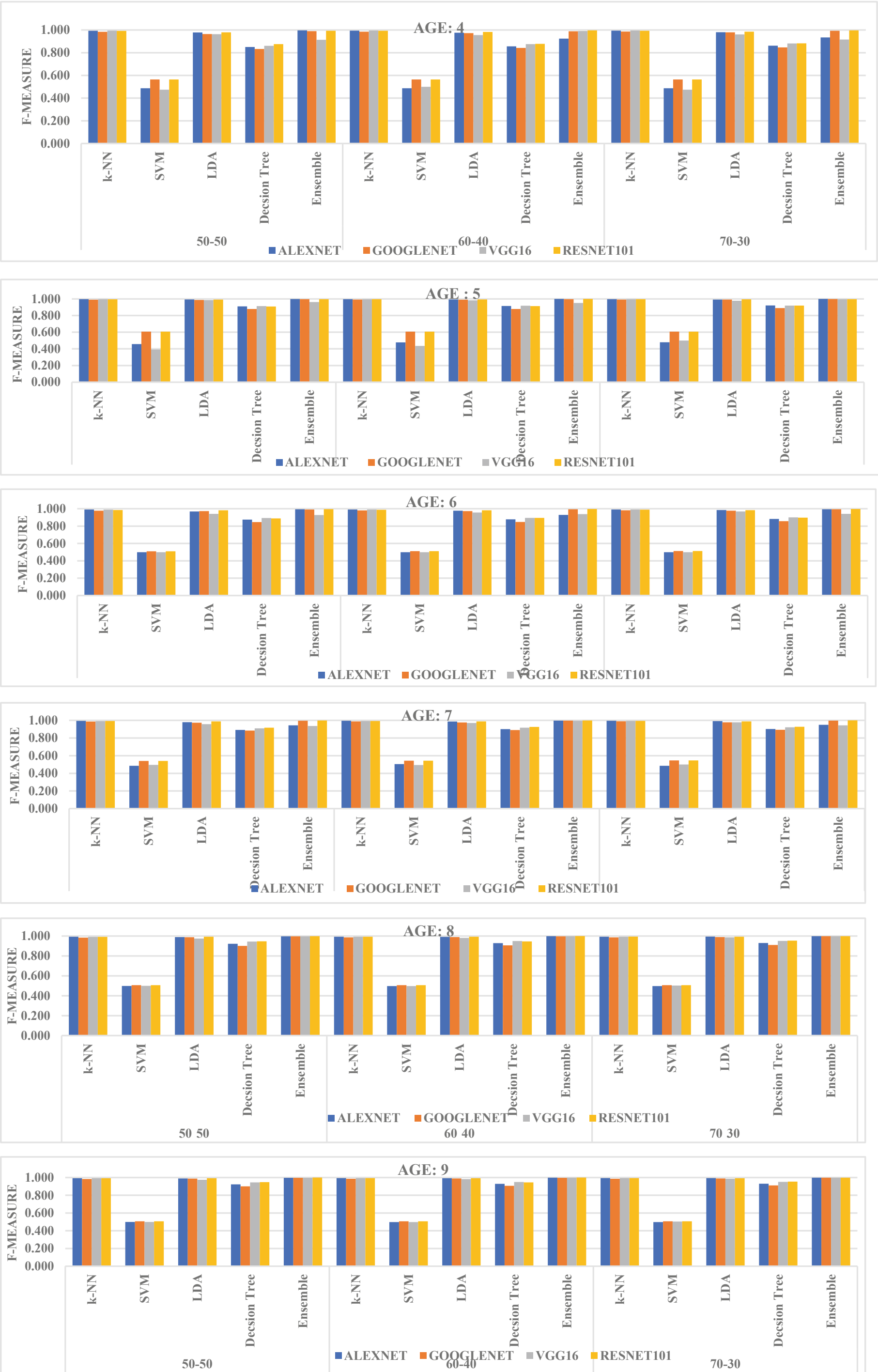

Fig. 6. F-Measure obtained from deep features for gender classification with age

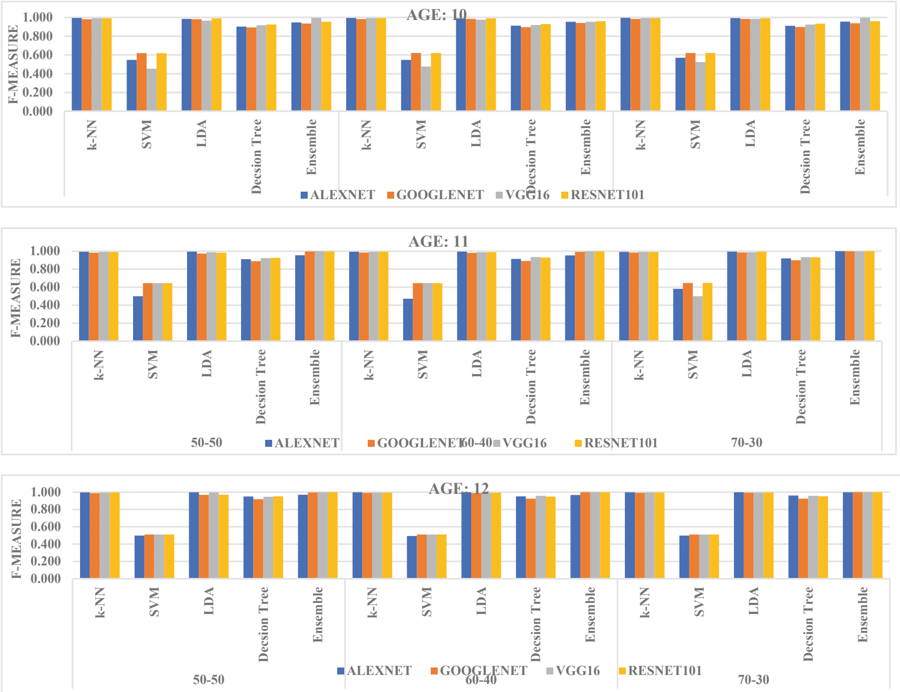

Fig. 6. (*continued*)

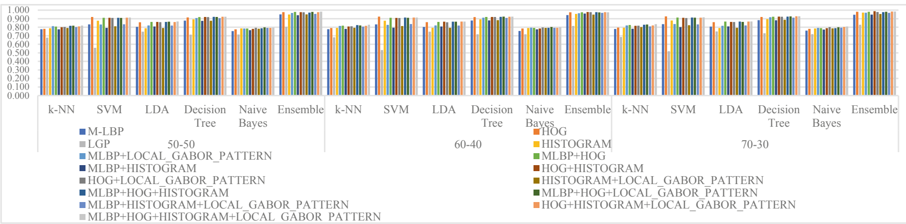

Fig. 7. F-Measure obtained from local features for gender classification without age information.

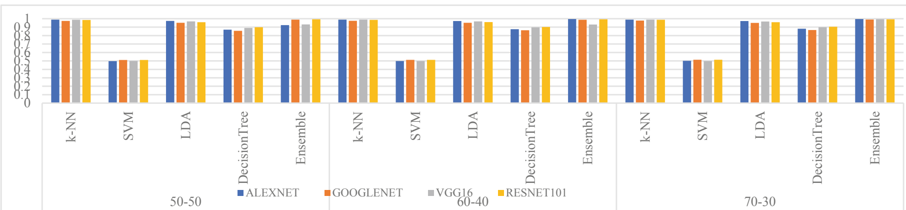

Fig. 8. F-Measure obtained from deep features for gender classification without age information.

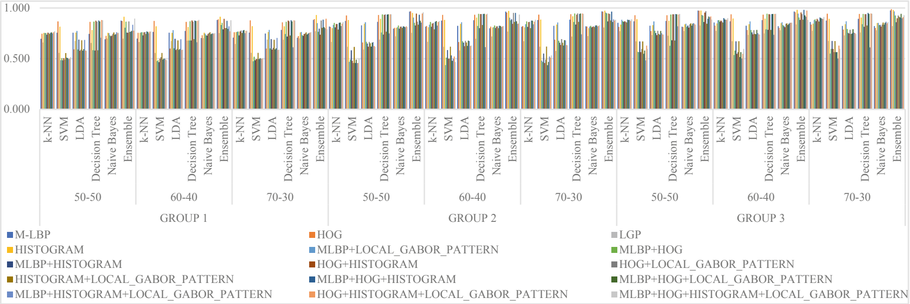

Fig. 9: F-Measure obtained from local features for group gender classification.

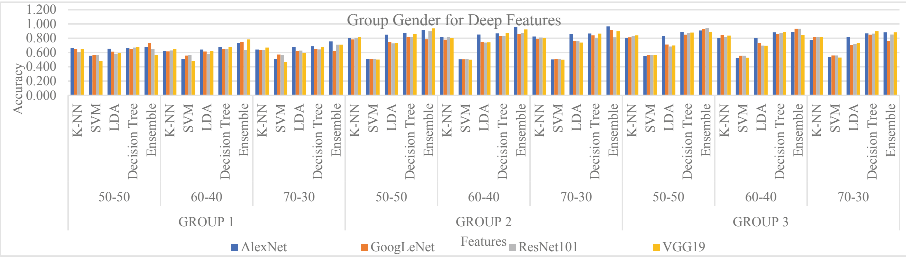

Fig. 10. F-Measure obtained from deep features for group gender classification.

5 Conclusion

In this work, we attempted children's gender classification problem by creating our own longitudinal face image dataset of very young children from age 4 to 14 years old. To address this problem, we have adopted two alternative approaches by varying feature extraction techniques: conventional and convolutional techniques. Post that classified on a multi-classification algorithm using ensemble technique. During the experimentation, the performance of gender classification with age information and without age information is evaluated. Finally, group-based gender classification is also computed for further analysis of our model. From this extensive experiment, we have observed the results for gender classification of young children using deep features are appreciated.

References

Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)

Balakrishnama, S., Ganapathiraju, A.: Linear discriminant analysis-a brief tutorial. Inst. Sig. Inform. Process. **18**(1998), 1–8 (1998)

Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)

Best-Rowden, L., Hoole, Y., Jain, A.: Automatic face recognition of newborns, infants, and toddlers: a longitudinal evaluation. In: Proceedings of the 2016 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE (2016)

Bledsoe, W.W.: Some results on multicategory pattern recognition. J. ACM **13**(2), 304–316 (1966)
Choi, S.E., et al.: Age estimation using a hierarchical classifier based on global and local facial features. Pattern Recogn. **44**(6), 1262–1281 (2011)
Craw, I., et al.: How should we represent faces for automatic recognition? IEEE Trans. Pattern Anal. Mach. Intell. **21**(8), 725–736 (1999)
Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1. IEEE (2005)
Friedl, M.A., Brodley, C.E.: Decision tree classification of land cover from remotely sensed data. Remote Sens. Environ. **61**(3), 399–409 (1997)
Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: SEXNET: a neural network identifies sex from human faces. In: NIPS, vol. 1 (1990)
Goodfellow, I.J., Shlens, J., Szegedy, C.: Explaining and harnessing adversarial examples. arXiv preprint arXiv:1412.6572 (2014)
Guo, G., et al.: A study on automatic age estimation using a large database. In: Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. IEEE (2009)
He, K., et al.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2016)
Keerthi, S.S., et al.: Improvements to Platt's SMO algorithm for SVM classifier design. Neural Comput. **13**(3), 637–649 (2001)
Keller, J.M., Gray, M.R., Givens, J.A.: A fuzzy k-nearest neighbor algorithm. IEEE Trans. Syst. Man Cybern. **4**, 580–585 (1985)
Krizhevsky, A., Hinton, G.:Learning multiple layers of features from tiny images, p. 7 (2009)
Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Adv. Neural. Inf. Process. Syst. **25**, 1097–1105 (2012)
Lawrence, S., et al.: Face recognition: a convolutional neural-network approach. IEEE Trans. Neural Netw. **8**(1), 98–113 (1997)
Lee. B, Gilani, S.Z., Hassan, G.M., Mian,A.: Facial gender classification — analysis using convolutional neural networks. In: Proceedings of the 2019 Digital Image Computing: Techniques and Applications (DICTA), pp. 1–8 (2019). https://doi.org/10.1109/DICTA47822.2019.8946109
Makinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned faces. IEEE Trans. Pattern Anal. Mach. Intell. **30**(3), 541–547 (2008)
Moghaddam, B., Yang, M.-H.: Learning gender with support faces. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 707–711 (2002)
Quinlan, J.R.: Bagging, boosting, and C4. 5. In: Aaai/iaai, vol. 1 (1996)
Rao, A., Noushath, S.: Subspace methods for face recognition. Comput. Sci. Rev. **4**(1), 1–17 (2010)
Safavi, S., Russell, M., Jancovic, P.: Automatic speaker, age-group and gender identification from children's speech. Comput. Speech Lang. **50**, 141–156 (2018). https://doi.org/10.1016/j.csl.2018.01.001
Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2014)
Toews, M., Arbel, T.: Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1567–1581 (2008)
Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society (1991)

Ullah, I., et al.: Gender recognition from face images with local WLD descriptor. In: Proceedings of the 2012 19th International Conference on Systems, Signals and Image Processing (IWSSIP). IEEE (2012)

Viola, P., Jones, M.: Robust real-time object detection. Int. J. Comput. Vision **4**(34–47), 4 (2001)

Wagner, S., et al.: Progression of gender dysphoria in children and adolescents: a longitudinal study. Pediatrics (2021)

Wang, H., et al.: Cosface: large margin cosine loss for deep face recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2018)

Wright, J., et al.: Robust face recognition via sparse representation. IEEE Trans. Pattern Anal. Mach. Intell. **31**(2), 210–227 (2008)

Yücesoy, E.: Speaker age and gender classification using GMM supervector and NAP channel compensation method. J. Ambient. Intell. Humaniz. Comput. (2020). https://doi.org/10.1007/s12652-020-02045-4

Zhang, W., et al.: Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. In: Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV 2005), vol. 1. IEEE (2005)

Zhuang, X., et al.: Face age estimation using patch-based hidden Markov model supervectors. In: Proceedings of the 2008 19th International Conference on Pattern Recognition. IEEE (2008)