International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Meta-Analysis of Multimodal Biometric Data for Age Classification using Deep Learning Techniques

Nagaraj T Bhoviwaddar

Rani Channamma University, Belagavi

A B S T R A C T

Age Classification of individuals is one of the most often conferred research work, and also it plays an important role in many application areas such as immigrant verification, health care and legal suits. In this work, Transfer Learning models have been applied for face and handwritten signature images to classify the Age trait of a person. A dataset of 7020 Face and Offline Handwritten Signature samples were collected from 351 adult volunteers (193 male and 158 female), age range between 18 to 60 years. We employed two transfer learning models XceptionNet and ResNet50, which are used for feature extraction as well as classification. The results of the proposed method outperformed good in fusion of dataset, that have shown a significant improvements recognition rate. The proposed framework multimodal system is also compared with other multimodal biometric system.

Keywords: Age Classification, Face, Offline Handwritten Signature, Transfer Learning Models

1. Introduction

Biometrics technologies are the new age systems, which allows an individual to be recognized and identified by their physiological and behavioral characteristics [1]. Recently, the technical advancement in the field of computer vision and pattern recognition shows the evolution in significance of human identity in immigrant authorization, legal and criminal investigations. Age estimation plays an imperative role in forensic and health sciences, where it determines the cases of unknown identity and suspected age groups [2]. Age is the demographic feature of a person which is related to physical characteristics like skin, weight, height and hair. Face and Handwritten signature data are universally, user accepted and convened biometric modalities. Due to high demand of accurate human authentication, modern technologies have surpassed the traditional authentication paradigms. However, in most of the cases using unimodal biometrics may fails to achieve recognition rate. In contrast to this, there are number of researchers who have worked on multimodal biometric data analysis are achieved satisfactory results. Multimodal biometric technology is a fusion of biometrics in terms of each unimodal feature fusion or integrating the scores or combining data. These are usually behaving better in accuracy, reliability and universality. In the recent era, deep learning-based models have shown a best performance in the image classification and pattern recognition. Estimating the age from the adults to adolescents is the significantly complex problem. Fundamentally, facial aging is not only assessed by genetic factors, it also depends on the lifestyle, expression and environment [3]. The same aged persons may look different in several locations, due to various reasons. Even in the case of Handwritten signature samples, each and all individuals have different way of writing skills. This modality is said to be cognitive behavioral biometric data and the assessment of age is based on the psycho-social development of person.

In the proposed work, multimodal biometric data for age classification is been presented to analyze the biometric data into respective age groups.

The rest of paper are followed by several sections: in section 2, existing related work on age classification is studied and defined the problem statement. Proposed methodology is been discussed in section 3 along with dataset description. Further, Experimental results and comparative results were discussed in section 4 and finally in section 5 Conclusion and feature work is been discussed.

2. Related Work

In the current state, many research works have provided insightful information for multimodal biometric data analysis. Age classification is a solemn field to develop mainly in forensic science and computer vision fields. In this section, several studies were listed below.

Vikas Sheoran et. al. [3] have proposed work on CNN based Age and Gender prediction system using transfer learning methods. In this study, UTKFace Dataset is used for age and gender classification. Three Deep Neural Networks or Transfer Learning Models like VGG16, ResNet50 and SENet50 were trained and tested. A highest accuracy for age classification 79.12% is obtained and for gender classification 94.94% is obtained using SENet50 model.

Dogucan Yaman et. al. [4] proposed the multimodal biometrics-based age and gender classification using ear and profile face images. Experiments are conducted on UND-F, UND-J2 and FERET datasets using two deep neural networks VGG16 and ResNet50 models. Fusion techniques were employed with center and SoftMax loss function. For unimodal biometrics 67.59% accuracy obtained on age classification and 99.11% for gender classification. For multimodal biometrics 67.59% accuracy for age classification and 99.79% for gender classification were obtained. In [5]. the same authors have proposed same framework using multi-task deep CNN for age and gender classification. Same datasets UND-F, UND-J2 and FERET datasets were used and same two DNN models VGG16 and ResNet50 models were used. Fusion techniques are also employed on each modality and results were analyzed. For Age classification 66.44% accuracy for combined data (ear + face) using ResNet50 and VGG16 have obtained 67.59% were obtained. For gender classification on combined data an accuracy of 99.11% obtained, using VGG16 and ResNet50 models.

Carmen Bisogni et. al. [6] proposed data fusion approach on periocular features of iris images to achieve age and gender classification. Extracted features were Blink, Fixation and Pupil of GANT dataset. Features are concatenated using feature level fusion techniques. Transformation and classifier-based score fusion approaches were employed i.e., Weighted sum, weighted product, and Bayesian rule. Three machine learning classifiers are used like gradient boosting, support vector machine and K-nearest neighbor. For gender classification, Combination of classifiers on weighted sum rule and weighted product rule have obtained 84.62%, and also using Bayesian rule 84.39% have obtained. For age analysis, weighted sum rule achieved 84.45%, product rule 84.45% and Bayesian rule on Random Forest and Support vector machine achieved an accuracy of 83.49%.

Maryem Erbilek et. al. [7] have proposed work on age prediction using multimodal biometric configurations. Two biometric datasets iris and handwritten signature were considered for experimentation. From signature static and dynamic signatures were extracted, while from the iris images geometric and texture features were extracted. These features were combined and trained using KNN, Jrip, Multilayer perceptron and SVM. Subsequently, all the feature combinations were considered and performed optimal mechanism to predict the age. The multimodal prediction system has achieved 90% accuracy, whereas unimodal biometric prediction system has achieved 75% accuracy.

In summary, studies have shown that the multimodal biometrics-based age classification will minimizes the search space and deals with various combination features and biometric data. This reflects age information could maximizes the recognition rate of the different biometrics data, that are more difficult to find.

3. Materials and Methods

3.1 Datasets

3.1.1 Face and Offline Handwritten Signature image samples

In this paper, we have used voluntarily own collected face and signature dataset consists of 7020 face and signature images with age and gender annotations. It has 3510 face images and 3510 signature images collected from 351 volunteers, in that 198 male and 153 female volunteers. The age range is from 18 years to 60 years.

- The Face images are collected using Cannon EOS1300D DSLR camera.
- The images cover with minimal variations of expressions, pose and illumination.
- Each participant was given knowledge about the purpose of collecting data.
- 10 samples were collected from each participant.
- The samples are shown in the figure 1.

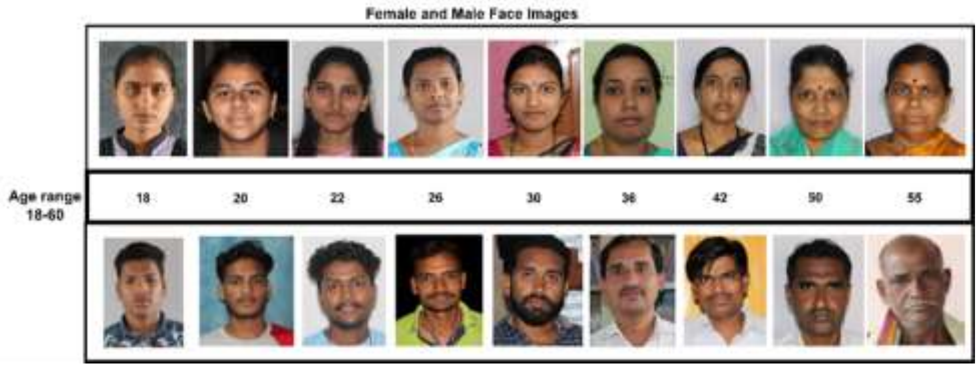


Figure 1. Face image samples

Whereas, Offline handwritten signatures are collected from same volunteers who all are participated in face image collections.

- Signature samples were collected on A4 white sheet, consists of predefined space for signatures.
- 10 samples were collected from each participant and their details of demographic information is also stored. Blue and black colored pens were used.

- The signatures are scanned using EPSON 1350D color scanner.
- All samples are cropped and used for the further analysis.
- The samples are shown in figure 2.

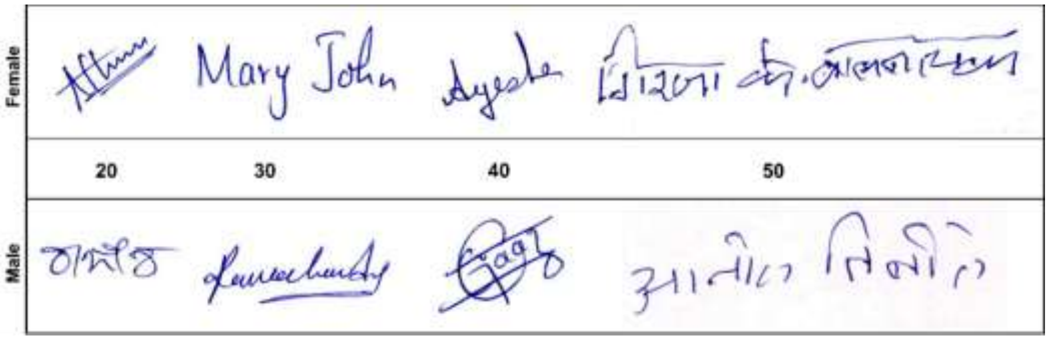


Figure 2. Offline Handwritten Signature Samples

4. Proposed Methodology

Classifying the age via extracting the face and handwritten signature images is the challenging task. The proposed framework is shown in the below figure 3.

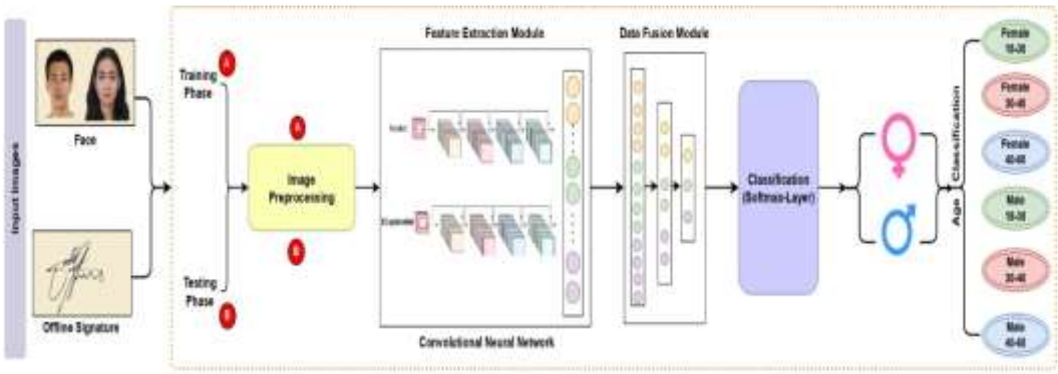


Figure 3. Illustration of the proposed framework

Pre-processing

Image Pre-processing is the fundamental step in the analysis, to get the accurate results: enhancement of the image is the important part. For the proposed work all collected image samples were resized into 224 x 224 x 3 image size and normalized to equal memory size.

Feature Extraction

Feature Extraction is major step in analyzing the images. Extracted features inhibits the distinguishable characteristics among the image samples. The proposed work utilized two transfer learning models for analysis i.e., ResNet50 and XceptionNet architectures.

•Fusion Techniques: The method of Concatenating more than one data is said to be fusion techniques. In the proposed work, we employed the data fusion method where two input data are combined and trained to check the robustness of the algorithm which classifies the images into respective age classes.

Transfer Learning Models

In the recent times, transfer learning models have produced more accurate and promising results in image classification and pattern recognition. Due to their distinctive feature map properties of images are directly extracted. The proposed work utilized two transfer learning model to improve the aforesaid problems. The following section explains the two models' description.

Xception Model

Xception was initially introduced by Francois Chollet in their paper entitled Xception: Deep Learning with Depthwise Separable Convolutions [8] as a convolutional neural network heavily influenced by the Inception architecture [9]. With the acronym standing for Extreme Inception, this architecture incorporates 36 convolutional layers designed for efficient feature extraction, organized into 14 modules. The key innovation of Xception lies in its utilization of depthwise and pointwise convolutions. By employing these techniques, Xception effectively reduces computational complexity and the number of parameters while still maintaining a high level of expressive power. The network structure of Xception is organized into three main parts: the entry flow, middle flow, and exit flow, as depicted in Figure 4. This organization allows for efficient learning and adaptability in a range of image classification and computer vision applications. The efficiency and adaptability of Xception make it a valuable tool for addressing various challenges in image classification and computer vision domains.

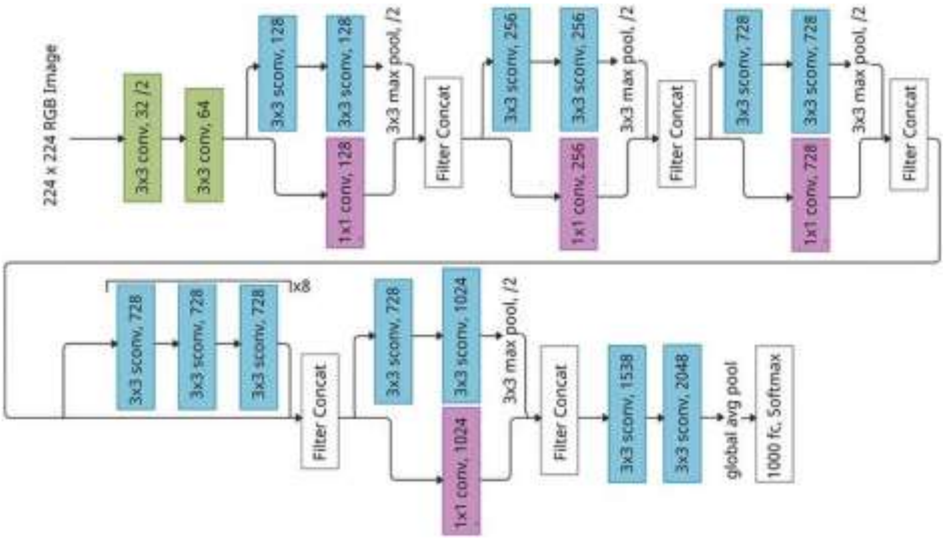


Figure 4. General architecture of Xception Model

ResNet50

ResNet50 was originally presented in 2016 by Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun in a seminal paper entitled Deep Residual Learning for Image Recognition [10]. This convolutional neural network (CNN) framework is classified under the ResNets family, which stands for Residual Networks. The key innovation brought forth by ResNet is the introduction of residual learning. In conventional deep networks, the addition of more layers can result in the vanishing gradient problem, wherein the gradients become excessively small, hindering the effective training of earlier layers. To overcome this obstacle, ResNet utilizes residual blocks that incorporate skip connections, also known as identity shortcuts. These connections enable the gradients to flow directly through the network. ResNet50, more specifically, denotes a ResNet architecture comprising 50 weight layers. It is recognized as a comparatively deep network capable of addressing a broad spectrum of computer vision tasks, including image classification, object detection, and segmentation. In Figure 5, the general architectural diagram for the ResNet50 model is depicted.

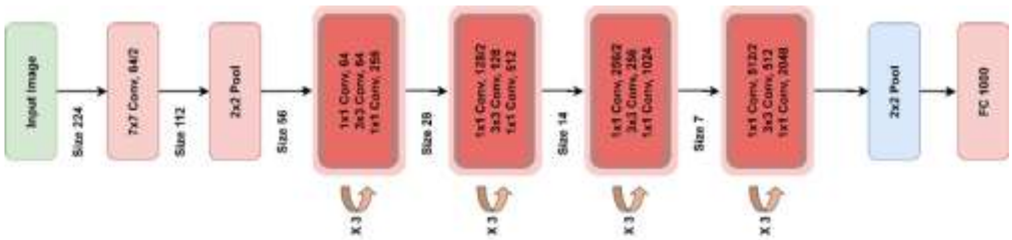


Figure 5. ResNet50 CNN Architecture diagram

5. Experimental Results and Discussion

In this section, we describe the experimental results, and a comparative analysis of the proposed work with existing work. We then illustrate the performance evaluation metrics used to assess the proposed work.

Implementation setup

We used Jupyter Notebook, a web-based interactive platform for Anaconda Python IDE, to implement the proposed work. We also used Tensorflow and Keras, Python libraries for deep learning, to analyze the data.

Results

In this study, two widely recognized deep neural network architectures, ResNet50 and XceptionNet, were employed to conduct age classification using face and handwritten signature biometric modalities. The dataset utilized in the experiment was divided into three portions: 70% for training, 20% for validation, and 10% for testing, as presented in Table 1. The pretrained network models, which had been trained on the ImageNet Dataset, served as the initialization for the networks. Subsequently, the networks were fine-tuned using the datasets in order to accurately classify gender. During the training phase, an initial learning rate of 0.001 was chosen, along with 20 epochs, and the Adam optimizer was employed. A batch size of 32 was selected for both unimodal and multimodal representation. The Max pooling technique and ReLU activation function were utilized in both networks, with a dropout rate of 0.5 employed after each fully-connected layer, excluding the output layer. For the final output layer, a SoftMax activation function was applied, along with a cross-entropy loss function, to prevent any performance deterioration in gender classification during the training phase.

Algorithmic Steps

The following algorithmic steps are used to perform Age classification from multimodal biometric images:

Input: Multimodal biometric images.

Output: Age classification from multimodal biometric images.

Step 1. Input the images from the dataset.

Step 2. Normalize the images into a size of 224x224x3.

Step 3. Extract features from the images using the ResNet50 and XceptionNet models.

Step 4. Fuse the input data from the two models and classify the images into age classes

Step 5. Train and test the feature map for validation.

Step 6. Evaluate the performance of the model on the test dataset.

Table 1: Data Split into ratio 70:20:10

Datasets	**Total**	**Train (70%)**	**Validation (20%)**	**Test (10%)**
Face Images	3510	2457	702	351
Offline Handwritten Signature Images	3510	2457	702	351
Combined Dataset	7020	4914	1404	702

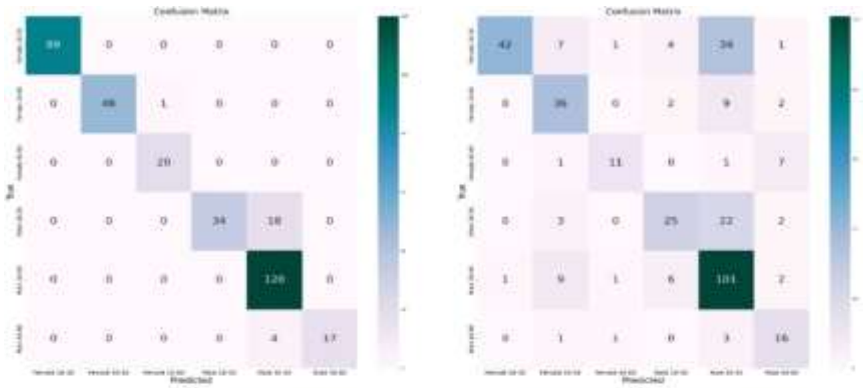
- **Age Classification Results**

The findings of the study regarding age classification using both unimodal and multimodal biometric techniques are presented in Tables 2. Additionally, the performance evaluation results can be found in Table 3. These results are based on two distinct datasets: face images and handwritten signatures. The experimentation process involved two main steps. Firstly, gender classification was attempted using individual models with the use of unimodal biometrics. secondly, both datasets were combined, and individual models were trained accordingly. Furthermore, the unimodal Face and Signature datasets yielded an age classification accuracy of 93.73% for face biometric and 74% for signature using ResNet50, and using XceptionNet 93.44% for face and 66% for signature. Lastly, the most accurate age classification results were achieved using a combination of the Face and signature dataset with the ResNet50, yielding an accuracy rate of 86.03%, and with XceptionNet yielded an accuracy of 80%.

The visual representation of the age classification results for the individual unimodal biometrics can be observed in Figure 6, Finally, Figure 7 illustrates the confusion matrix for the combined dataset using both models.

Table 2. Age classification results using different modalities

Datasets	**XceptionNet**	**ResNet50**
	Test Accuracy	**Test Accuracy**
Face	93.44%	93.73%
Signature	66%	74%
Face + Signature	80%	86.03%



a) Face biometric with XceptionNet b) Handwritten Signature biometric with XceptionNet

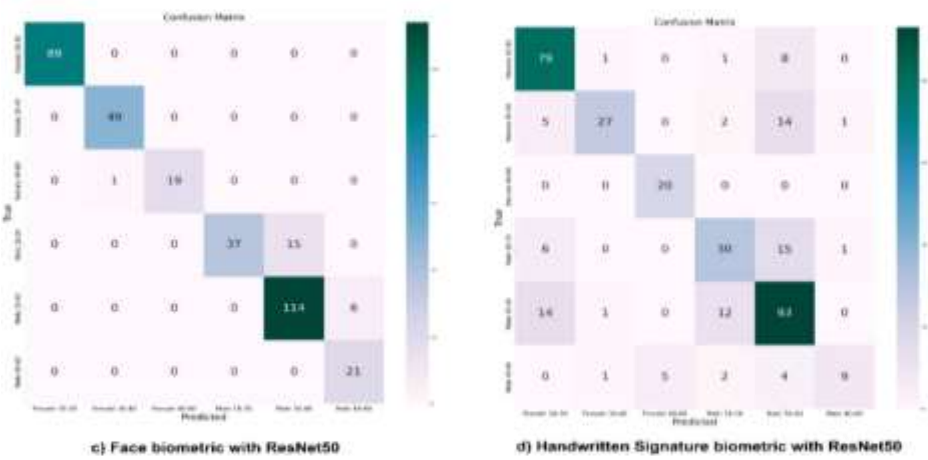


Figure 6. Confusion Matrix of test data results of Face and Handwritten Signature Dataset

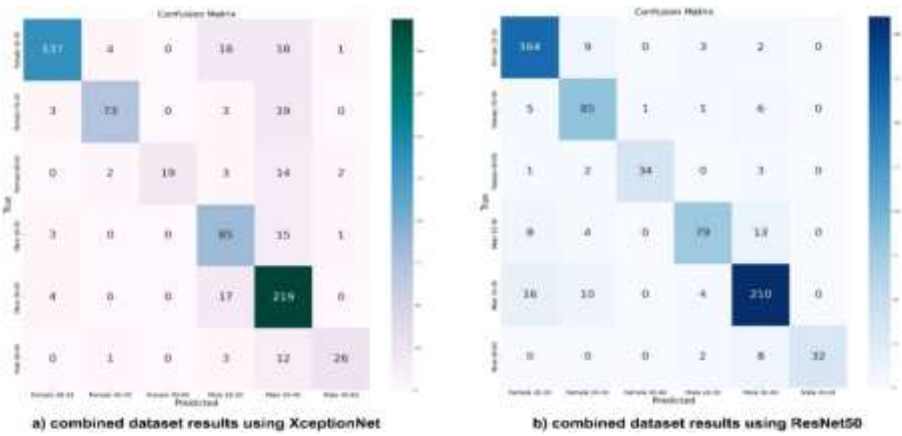


Figure 7. Confusion Matrix of test data results of combined Dataset

Performance Evaluation

Performance evaluation results were examined in order to assess the effectiveness of the proposed model, particularly in relation to the metrics achieved on the designated dataset. The metrics employed in this investigation are explicitly stated below.

- **Accuracy:** is determined using the proportion of correct values to the total number of values.

$$Accuracy = \frac{No.of\ correctly\ predicted\ values}{total\ no.of\ values} \qquad (1)$$

The terms used in the performance metrics are given below;

- True Positive (TP) means the prediction was correct.
- True Negative (TN) means the prediction was incorrect.
- False Positive (FP) means the prediction was correct but it shouldn't have been.
- False Negative (FN) means the prediction was incorrect but it should have been.
- **Precision:** This metric determines the accuracy constraints. It can be measured as the proportion of true positive predictions to total positive predictions.

$$Precision = \frac{TP}{(TP+FP)} \qquad (2)$$

- **Recall:** Recall focuses on determining the proportion of real positives that were misidentified. The following formula can be used to calculate recall**:**

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

- **F1-Scores:** F1-Scores are a form of composite score that includes both precision and recall. As a result, the F1 score can be calculated as the harmonic mean of precision and recall, with equal weight given to each.

$$\boldsymbol{F1 - Scores} = 2 * \frac{precision * recall}{precision + recall} \quad (4)$$

Table 3: Performance evaluation results of the proposed method

Models	*Datasets*	*Precision*	*Recall*	*F1-Scores*	*Accuracy*
XceptionNet	**Face**	0.94	0.91	0.93	**93.44%**
	Handwritten Signature	0.72	0.66	0.65	66%
	Face + signature	0.85	0.80	0.80	**80%**
ResNet50	**Face**	0.94	0.94	0.93	**93.73%**
	Handwritten Signature	0.77	0.74	0.73	74%
	Face + signature	0.89	0.86	0.86	**86%**

In inference, the results of the data fusion experiment demonstrate a statistically significant improvement in performance over a monomodal system.

Comparative Analysis

Table 4 provides a comparison of existing and proposed methods for gender classification. It summarizes the most recent findings, feature extraction methods, classification approaches, and results obtained using multimodal biometric data.

Table 4: Overview Comparative analysis on Gender Classification Systems

Author	**Methodology**	**Dataset**	**Classifiers**	**Results**
Vikas Sheoran et. al. [4]	VGG16, SENet50	UTKFace dataset	VGG16, SENet50	79.12% 94.94%
Dogucan Yaman et. al [5]	ResNet-50 VGG16	FERET, UND-F and UND-J2 datasets	VGG16 and ResNet-50	67.59% for combined both models
Dogucan Yaman et. al [6]	ResNet-50 VGG16	FERET, UND-F and UND-J2 datasets	VGG16 and ResNet-50	66.44% for combined both models
Carmen Bisogni et. al. [7]	Blink, Fixation and Pupil features of iris images	GANT dataset	SVM, RF, KNN	84.62%, 84.45%, 83.49%
Maryam Eskandari and Omid Sharifi [19]	Uniform LBP and Backtracking search algorithm (BSA)	CASIA-Iris and MBGC multimodal biometrics	Support Vector Machine	94%
Proposed Method	**XceptionNet** **ResNet50**	**3510 Face images, 3510 Offline Signature images, 7020 Face +Offline Signature images**	**XceptionNet** **ResNet50**	**80% XceptionNet, 86% ResNet50 for both dataset**

6. Conclusion

This paper describes a study that uses deep learning approaches to provide a multimodal biometric data analysis technique for age classification. Security, document verification, and criminal investigations all rely heavily on age classification systems. Our technique aims to appropriately classify an individual's data by taking into account both physical and behavioral biometrics. To do this, we use two transfer learning approaches for domain adaptation: the ResNet50 and XceptionNet models. We use a fusion technique to merge data from two independent datasets to improve the distinctiveness of the features. Following that, we examine each biometric data set using the aforementioned models and obtain encouraging findings. Furthermore, our method investigates the role of profile face photos and offline handwritten signature images in age classification.

Conflict of Interest

There is no conflict of interest declared by authors.

References

[1] Gornale, S. S., Kumar, S., Patil, A., & Hiremath, P. S. (2021). Behavioral Biometric Data Analysis for Gender Classification Using Feature Fusion and Machine Learning. Frontiers in Robotics and AI, 8. https://doi.org/10.3389/frobt.2021.685966

[2] Baydogan, M. P., Baybars, S. C., & Tuncer, S. A. (2023). Age-Net: An Advanced Hybrid Deep Learning Model for Age Estimation Using Orthopantomograph Images. Traitement Du Signal, 40(4), 1553–1563. https://doi.org/10.18280/ts.400423

[3] Sheoran, V., Joshi, S., & Bhayani, T. R. (2021). Age and Gender Prediction Using Deep CNNs and Transfer Learning. In Communications in Computer and Information Science (Vol. 1377 CCIS, pp. 293–304). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-981-16-1092-9_25

[4] Yaman, D., Eyiokur, F. I., & Ekenel, H. K. (2019). Multimodal age and gender classification using ear and profile face images. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (Vol. 2019-June, pp. 2414–2421). IEEE Computer Society. https://doi.org/10.1109/CVPRW.2019.00296

[5] Yaman, D., Eyiokur, F. I., & Ekenel, H. K. (2022). Multimodal soft biometrics: combining ear and face biometrics for age and gender classification. Multimedia Tools and Applications, 81(16), 22695–22713. https://doi.org/10.1007/s11042-021-10630-8

[6] Bisogni, C., Cascone, L., & Narducci, F. (2022). Periocular Data Fusion for Age and Gender Classification. Journal of Imaging, 8(11). https://doi.org/10.3390/jimaging8110307

[7] Erbilek, M., Fairhurst, M., & Da Costa-Abreu, M. (2014). Improved age prediction from biometric data using multimodal configurations. In Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI) (Vol. P-230, pp. 179–186). Gesellschaft fur Informatik (GI).

[8] Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. In Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 (Vol. 2017-January, pp. 1800–1807). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/CVPR.2017.195

[9] Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Wu, X., Li, X., & Wang, Q. (2021). DeepCervix: A deep learning-based framework for the classification of cervical cells using hybrid deep feature fusion techniques. Computers in Biology and Medicine, 136. https://doi.org/10.1016/j.compbiomed.2021.104649

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Vol. 2016-December, pp. 770–778). IEEE Computer Society. https://doi.org/10.1109/CVPR.2016.90