REVIEW

Review of age and gender detection methods based on handwriting analysis

Fahimeh Alaei[1] · Alireza Alaei[1]

Abstract

Handwriting recognition and analysis has been an active area of research in the last two decades. Handwriting analysis is being studied in various fields of science, such as graphology, neurology, psychology, and computer science. Furthermore, automated handwriting analysis has several applications, including forensic, security, medical, and disease prediction. This paper presents the most recent handwriting analysis techniques and advancements available in the literature for age and gender classification/detection. Different steps, including feature extraction and classification, frequently used in the literature for age and gender detection, are discussed, and the presented works are classified according to the applied feature extraction and classification methods. The online and offline benchmark databases are also reviewed. We used a text mining technique to perform a quantitative content analysis of the presented research and better understand the co-occurrence network diagrams of age and gender classification/detection. This study is a valuable resource that provides new research directions to students and researchers interested in this field for further research and investigation.

Keywords Handwriting analysis · Age classification · Gender detection · Text mining

1 Introduction

Handwriting is a multi-sensory activity and skill that plays a vital role in everyday life. This skill is called brain writing, which defines the reflection (mood, nature, and personality) and complex motor skills (age, gender, and nationality) of an individual brain [1, 2]. An individual generally creates a unique handwriting pattern. This uniqueness has encouraged researchers to use handwriting, as a behavior biometric, for several tasks, including forensics, security, and disease prediction. Handwriting analysis or graphology can help identify, assess, and interpret individuals' behavior using their handwritten samples. Graphology uses the characteristic of personality and emotional activities for various applications such as writer verification/identification, handwriting recognition, signature verification, age and gender classification/

✉ Fahimeh Alaei
   fahimeh.alaei@scu.edu.au

   Alireza Alaei
   ali.alaei@scu.edu.au

[1] Southern Cross University, Gold Coast, Australia

detection, and disease prediction. Automated handwriting analysis can improve the performance of these applications. Thus, developing automated handwriting analysis has become one of the active research areas in the last few decades to understand and identify an individual's personality automatically.

Detecting an individual based on handwriting is a kind of behavior biometric identification well-acknowledged by psychologists, neurologists, paleographers, forensic analysts, document analysts, and computer science researchers. Similarly, the existence of a relationship between handwriting and different demographic attributes of writers, such as gender, handedness, and age, is also confirmed by psychologists and neurologists [2–6]. For instance, it has been reported that females' handwritings are generally neat, delicate, consistent, regular, homogenous, attractive, and decorative. In contrast, male handwriting tends to be hurried, untidy, scruffy, and spiky [2, 6]. Gender detection of the writer from a handwritten document has been a challenging problem for the document analysis community [2, 6].

From the handwriting analysis perspective, gender detection is a task between handwriting recognition and

writer identification. In handwriting recognition, the variability of handwriting should be blurred to raise a common feature among the writers to enable more accurate handwritten document recognition. In contrast, the variability in the writer identification task should be highlighted to increase differences among the handwritings of different writers in order to achieve more accurate gender/age classification results. In gender detection, however, we look for common features between a particular group (male or female) of writers [7]. The gender detection problem is a binary classification, while age detection is a multi-class classification problem. In general, to address the problems of gender and age detection, signal and image processing followed by pattern classification techniques are applied to handwriting signals or images. The methods in the literature are grouped into online and offline categories according to data acquisition. The handwriting samples written on paper with a typical pen/pencil and converted into digital documents using a digital camera or scanner are categorized as offline samples. Unlike offline, online handwriting samples are collected digitally, using the iPad/tablet and digital pen [8, 9].

Given online and offline categories, different types of features may be extracted from handwriting documents. For instance, in online age and gender detection, features such as time-in-air, time-on-surface, velocity, acceleration, pen pressure, x–y coordinates of the pen position, writing trajectory, and order of strokes, carry important information about the age and gender of the writer. In contrast, storing this information for the offline category or extracting them from document images is impossible. Features extracted from online or offline handwriting samples can be divided into macro and micro [10]. The macro features describe the overall pictorial characteristics (size, slant, shape, and space between the words and characters) of handwriting samples, whereas micro features describe the attributes of individual characters/components, for example, the geometry or shape of the individual characters. The macro features were then used to find the similarity or distance measure between the two documents. However, micro features were used to compute the similarity of characters and correlation measures [11].

In addition, features used for characterizing handwriting can be categorized into two groups, conventional handcrafted and deep learning-based features, based on the types of methodologies used in the literature. Examples of handcrafted features are textural and structural features that can be extracted from the word, line, and whole documents. Deep learning features are those extracted from different layers of, for example, a convolutional neural network (CNN) based model. It is worth noting that a limited number of handwriting samples can be collected from writers, resulting in small benchmark databases, making extracting deep learning features challenging.

Considering the dependency of age and gender classification approaches on scripts (the language of writing), two categories of methods, script-dependent and independent, were proposed in the literature. Script-dependent refers to selecting training and testing samples from the same scripts. In contrast, when training and testing samples are selected from different scripts, for example, training samples are English manuscripts, and testing samples are Arabic manuscripts, the approach is called script-independent.

Various research has been conducted on age and gender detection using handwriting. A comprehensive survey for age and gender detection in recent years can provide readers with a summary of research work and the development in this research area. Thus, this article reviews recent papers about age and gender classification/detection in the literature. We believe this study would be helpful for novice researchers in the handwriting analysis field, in general, and in gender and age detection techniques, in particular, and provide researchers with insights and new research directions. Overall, this article contains the following contributions and tries to address the following points:

- discussion of the trend of age and gender classification/detection systems in the last decade using a text mining technique;
- comparison of the results obtained from traditional and deep learning techniques and providing readers with information about the merits and demerits of methods and identifying methods that obtained the highest accuracies for age and gender detection so far;
- identification of the feature extraction and classification methods performed better on the benchmark databases for gender and age detection tasks;
- a detailed study of commonly used databases in the literature for age and gender detection, their scripts, the number of writers, and the number of samples.

The rest of this article is organized as follows. Section 2 provides the methodology of the research. An overview of a general age and gender detection framework is discussed in Sect. 3. The related works in this domain are discussed in Sect. 4. Metrics and databases used for experimentation and comparison analysis are presented in Sect. 5. Section 6 provides discussion and remarks. Finally, the conclusion and future directions are given in Sect. 7.

2 Methodology

The methodology applied to collect, filter and finally choose relevant papers for review is depicted in Fig. 1. As shown in Fig. 1, we used different databases, including Google Scholar, Web of Science, ProQuest, and EBSCO, to download relevant peer-reviewed papers. The databases were carefully selected based on their privileges, such as credibility, reliability, availability of peer-reviewed content sources, and advanced search function. Several keywords, such as "gender detection using handwriting", "gender classification using handwriting", "gender identification using handwriting", "age classification using handwriting", and "age identification", were used for searching the databases. As a result, 103 papers were downloaded and collected accordingly.

Zotero, a reference management software, was used to manage and work with the downloaded papers. The software allows for storing and organizing research papers and sharing references. After checking for duplication, 35 papers were eliminated. Since our research focused on the studies published from 2012 to 2023, 12 more papers were further eliminated from the list of collected papers. In the subsequent filtering, 14 more papers were eliminated as they were not published in high-ranked journals and conferences. The remaining 42 papers were finally considered for this study. Considering the dates of publications, we noted that the number of publications on gender detection was more consistent in contrast to the age classification. Most research publications were conducted between 2014 and 2017 for both age and gender detection. There was a gap and lack of focus on the age classification using handwriting between 2016 and 2021 and a sudden rise in 2022. Although there was a fluctuation in the number of publications on gender detection from 2012 to 2023, there was no significant gap in publishing papers in this area. This could be because of various gender detection applications and the availability of databases and metadata in the literature.

3 Overview of general age and gender detection

Developing an automated handwriting analysis system to detect a gender or age category from handwriting samples involves two stages, developing and training a model and then testing the trained model. A block diagram of a general system for age and gender classification/detection is represented in Fig. 2. As demonstrated in Fig. 2, similar pre-processing and feature extraction methods are employed in both the training and testing stages. Pe-processing techniques are applied to improve handwriting quality as a preliminary step. These techniques may be used in different sequences, such as binarization, segmentation, and normalization. For instance, the pre-processing step may segment a handwritten document into lines, words, characters, or patches.

Feature extraction comes right after the pre-processing step to compute a set of informative features and to support training in the classifier step. Features may be extracted from different levels of a sample, such as a pixel, patch, or whole sample. In addition, features may be extracted using handcrafted or deep learning methods. Various classifiers,
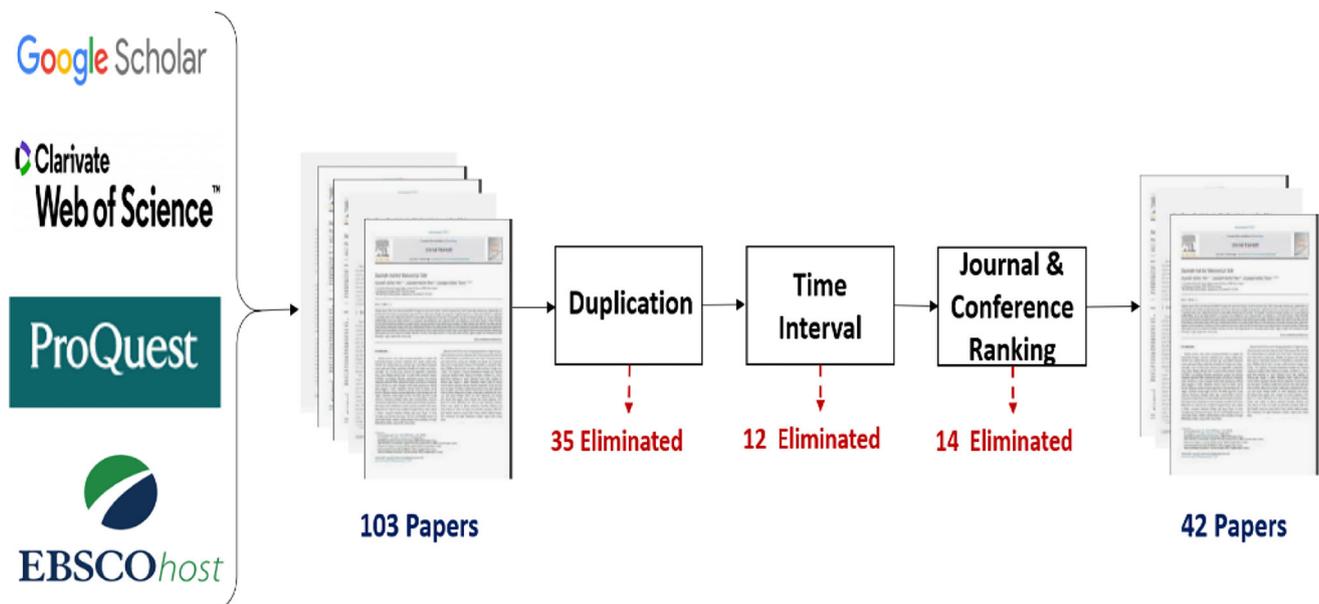

Fig. 1 The methodology used for the collection and selection of the papers
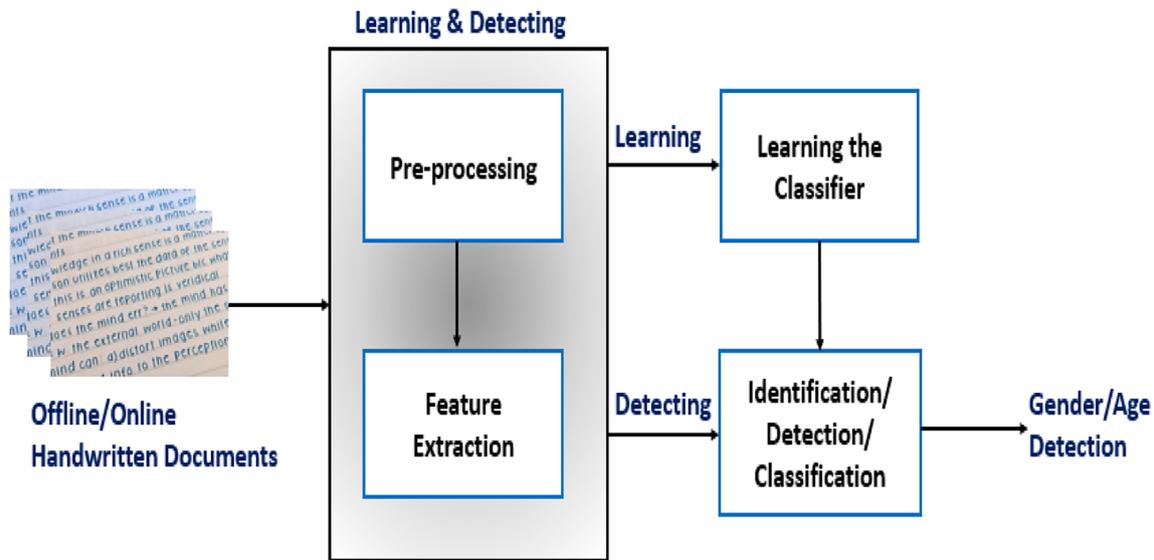
Fig. 2 Block diagram of a general framework for age and gender classification/detection

such as Support Vector Machine (SVM), logistic regression, K-nearest neighbor (K-NN), decision tree (DT), random forest (RF), artificial neural network (ANN), and deep neural network, may be trained for gender/age detection or classification.

In the testing stage, similar to the training, the pre-processing and feature extraction techniques are applied to a new sample. The extracted features are passed to the trained classifier to carry out the age and gender classification/detection task.

4 Related works

Demographic attributes of writers, such as age or gender, were considered in several papers in the literature for handwriting analysis [8, 12, 13]. Some papers further considered a combination of demographic attributes of writers for analysis, for example, gender and handedness of writers, to improve the performance of automated handwriting systems in the literature [14, 15]. Various approaches employed in the literature for age and gender classification/detection through handwriting are reviewed in this section. Figure 3 illustrates a categorization of different approaches in age and gender classification/detection research, including traditional and deep learning. Each category is further divided into offline and online methods based on data acquisition types. It is noted that there is no significant amount of research on age classification available in the literature compared to gender detection. The studies that characterized only the writer's age are summarized in the age classification sub-section. The studies that characterized the writer's gender and a combination of

the writer's gender and age as demographic attributes are discussed in the gender detection sub-section.

4.1 Age classification

As mentioned earlier, only a few research studies focused on the age classification problem [12, 16–18]. The age classification methods in the literature can be grouped into offline [12, 16], and online [17, 18], considering the input data used for analysis.

Traditional and deep learning methods were considered in the offline age classification approaches. Disconnectedness features extracted from Canny and Sobel edge images, and k-means clustering were used for a four-class age classification problem in [12]. The proposed method was applied on a database, and an accuracy of 66.25% was obtained. The IAM and KHATT databases were also used to evaluate the proposed methods, and age classification accuracies of 63.6% and 64.4% were obtained, respectively. Considering the deep learning approach for age detection, ResNet and GoogleNet were used for transfer and feature learning [16]. An SVM classifier was then used for age detection [16]. The proposed method was tested on the FSHS database. Considering ResNet and GoogleNet features, 69.7% and 61.1% correct age detection were obtained from the proposed method [16]. The same authors further considered different features and SVM and NN for age detection in two age categories, youth-adult and mature-adult, and accuracies of 71% and 63.5% were obtained, respectively [19]. However, the main issue with deep learning based methods, for example, ResNet, is that when a network becomes deeper, it will be more challenging to train it effectively due to issues such as
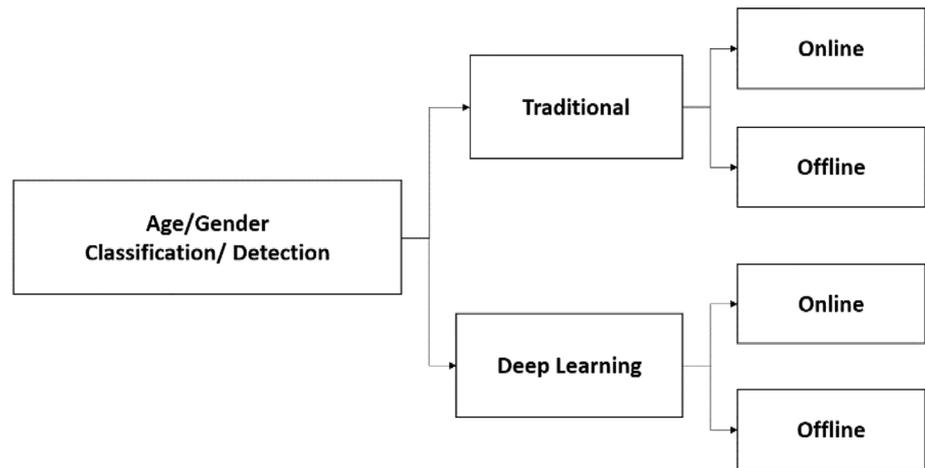
vanishing or exploding gradients. In addition, the performance of deep learning based methods relies heavily on the quality and diversity of the training dataset, and collecting high-quality data for age detection is difficult due to privacy and human factors. It is worth noting that the SVM classifier was originally developed for binary classification and regression. It was further modified for multi-class problems and showed its superiority in various applications. The basic idea behind SVM is to find the best possible decision boundary, called a hyperplane, that separates different classes of data points. The SVM algorithm tries to maximize the margin, which is the distance between the hyperplane and the nearest data points of each class. The intuition is that a larger margin indicates a better generalization ability of the classifier.

Age characterization from online handwriting was examined using supervised and unsupervised learning approaches in the literature [18]. The authors used a 2-level clustering approach for online handwriting characterization style. In the first level, raw spatial-dynamic information was extracted, whereas, in the second level, the style of word variability was extracted and converted into a Bag of Prototype Words (BPW). At the first level, Linear Discriminant Analysis (LDA) detected the features that distinguish different age groups. For instance, teenagers' handwriting has shown the highest stability compared to elders' handwriting, although there are no significant differences in the middle-aged group. The K-means clustering was considered for age detection [18]. However, the LDA based approach is unimodal and may not be applicable in different scenarios. The Norm Discriminant Eigenspace Transform or Moments discriminant analysis may be utilized to address this issue. The same authors further discussed several patterns, such as high time on air, slow velocity, acceleration, the highest number of smallest strokes, and less fluent handwriting, in people's handwriting by comparing the elders' handwriting with middle-aged

handwriting [8]. The sequential forward floating selection (SFFS) method was also considered for classifying children from adult handwriting [17]. Online handwriting text and patterns were collected from several children and adults. Different features, such as pen pressure, time of writing, and pen angle, were extracted from handwriting samples and then selected using the SFFS method. SVM and RF methods were applied for classifying the samples, and the age detection accuracies of 87.4% and 91.5% were obtained from the handwritten text, respectively.

A summary of age detection methods in the literature is provided in Table 1. From Table 1, it can be noted that online methods performed better than offline methods in age classification. When looking at the results presented in Table 1, it is important to note that the direct comparison of the results obtained from various methods applied on different databases is not fair; however, the implication and significance of the features and signals used in different handwriting analysis methods can be considered as an important finding.

4.2 Gender detection

The proposed gender detection methods in the literature are designed based on traditional and deep learning approaches to deal with offline and online handwriting samples. Details of traditional and deep learning approaches and the analysis of keywords used in the gender detection literature are discussed in the subsequent subsections.

4.2.1 Traditional methods

In most of the conducted research in the gender detection area, traditional methods were applied to offline document images [6, 13, 20]. Traditional methods are generally designed based on handcrafted features extracted from handwriting images, followed by the classification using

Table 1 Overview of the recent research on age classification using traditional and deep learning methods

Method	Study	Database	Script	Setup	Feature	Classifier	Accuracy %
Offline	[12]	No official name	English	S = 400, C = 3	Disconnectedness features	K-means	66.25
			Arabic				63.60
		IAM		S = 168, C = 2			64.44
		KHATT		S = 838, C = 2			
	[16]	FSHS	Arabic	S = 2000, C = 2	ResNet,	SVM	69.7
					GoogleNet		61.1
Online	[17]	No official name	Japanese	S = 1710, C = 2	Pen pressure,	SVM	87.4
					Time of writing,	RF	91.5
					Pen angle		
	[18]	IRONOF	English French	W = 144, C = 6	Bag of prototype words, linear discriminant analysis	K-means	–

(S is the number of samples, W is the number of writers and C is the number of age categories in experimental setups)

one or a set of classifiers. Various feature extraction methods based on codebook, geometric, texture, or a combination of both geometric and texture were investigated in the gender detection literature [6, 13, 20]. The geometrical features were used to measure the orientations, curvatures, roundness, slants and strokes of the characters. Texture features, however, were used to extract textural characteristics from words, patches, or whole images. These characteristics were used to generate a histogram of features.

A summary of the traditional approaches in the literature that are ordered based on the databases used for evaluation is presented in Table 2. From Table 2, it can be noted that most of the available databases, metadata, and methods in the literature were developed for offline approaches [6, 13, 20, 21]. Each method presented in Table 2 is detailed in the following.

Authors in [22] used geometrical and transformed features followed by kernel mutual information consisting of the kernel function as a feature selection technique for gender classification. The classification was carried out using SVM on the Chinese Registration Document Form (RDF) database and the ICDAR 2013 database consisting of English and Arabic document images. Accuracies of 66.7% and 66.3% were obtained from their proposed method on RDF and ICDAR 2013 databases, respectively. The same authors further discussed the minimal-Redundancy-Maximal-Relevance (mRMR) based on the mutual information method for selecting the features for the gender identification task [23]. The geometrical features, such as orientation, roundness, slant, and transformed features, including the Gabor and Fourier features, were also considered in this study. Different shape descriptors, such as

tangent angle function, curvature function, and Fourier descriptors, were also considered for gender detection in Bosnian document images, BHDH a database consisting of 3766 Bosnian document images [24]. Notably, the mRMR captures linear relationships between variables and cannot capture complex nonlinear relationships between features. Unlike mRMR, Recursive Feature Elimination (RFE) concentrates on the relevance of features to the target variable and assigns weights or ranks to features based on their importance in predicting the target variable and can be used for feature transformation.

In [25], curvatures, directions, tortuosity, and chain codes as a set of geometric features were further considered to characterize age, gender and nationality in handwritten document images. The kernel discriminant and random forests classifiers were applied to perform age, gender, and nationality prediction. The proposed systems were evaluated using the QUWI database, and the chain code feature, with an accuracy of 74.05%, provided better gender prediction results than other features. To address the problem of using only local features for document characterization, Siddiqi et al. [26] proposed local and global features, including curvature, slant, and legibility, for gender prediction [26]. They then used SVM and ANN classifiers to detect gender from handwritten documents. They finally evaluated their proposed method using QUWI and MSHD databases, and correct gender detections of 68.75% and 73.02% from QUWI were obtained from the MSHD database [26]. Ibrahim et al. also used local and global features (gradient and wavelet domain local binary patterns (WD-LBP)) along with an SVM classifier for gender identification [27, 28]. The experimental results were conducted using the ICDAR 2013 dataset, and the highest

Table 2 Overview of the recent research on gender detection using traditional methods

Study	Database	Script	Setup	Feature	Classifier	Accuracy (%)
[25]	QUWI	English Arabic	S = 4068, TR = 70%, TS = 30%	Geometric features	RF, Kernel discriminant	74.05
[30]	QUWI	English and Arabic	TR = 300, VL = 100, TS = 100	LBP, HOG, GLCM	SVM, ANN, DT, RF	85.00
[13]	QUWI	English and Arabic	TR = 300, VL = 100, TS = 100	Gabor filter, Fourier transform	NN	70.00
[7]	QUWI	English and Arabic	TR = 500, VL = 250, TS = 250	Cloud of Line Distribution, Hinge feature	SVM	73.60
[6]	QUWI ICDAR 2013, ICDAR 2015, ICFHR 2016	English and Arabic	TR = 282, TS = 193 TR = 300, VL = 100, TS = 100 TR = 500, VL = 250, TS = 250	oBIFs	SVM	71.00, 76.00, 68.00
[26]	QUWI MSHD	English and Arabic French and Arabic	TR = 1200, TS = 400 TR = 252, TS = 252	Slant, Curvature, Neatness, Fractal features, Texture features	ANN SVM	68.75 73.02
[21]	QUWI MSHD	English and Arabic French and Arabic	TR = 282, TS = 193 TR = 252, TS = 252	Wavelet transform, Probabilistic Finite state automata	ANN SVM	77.70 77.60
[39]	QUWI MSHD	Arabic and English English and French	TR = 300, VL = 100, TS = 100 TR = 950, TS = 950	SIFT descriptor	SVM Ensemble classifier	82.00 90.00
[29]	IAM	English	TR = 100, TS = 50	Curvelet transform	OC-SVM	77.33
[31]	IAM	English	TR = 200, VL = 50, TS = 100 TR = 80, VL = 20, TS = 50	LBP HOG	SVM	70.00 74.00
[14]	IAM KHATT	English Arabic	TR = 80, VL = 20, TS = 50 TR = 180, TS = 90	HOG GLBP	SVM	70.00
[27]	ICDAR 2013	English Arabic	S = 1900	Gradient, WD-LBP	SVM	94.70
[23]	ICDAR 2013 RDF	English and Arabic Chinese	S = 1900	Geometrical, Transformed features	SVM	66.30 66.70
[40]	FSHS	Arabic	TR = 1400, VL = 300, TS = 300	Space between words, Handwriting irregularity, Margins	ANN SVM	97.10 94.70
[41]	BiosecurID	Spanish	S = 242	Pen-down, Pen-up strokes	Distortion	74.00
[20]	No official name	Gurumukhi	W = 200	Curve fitting-based features	Hybridization	90.57
[35]	No official name	–	TR = 60, TS = 15	Height, pressure, Margin	Chi's Algorithm	76.00

Table 2 (continued)

Study	Database	Script	Setup	Feature	Classifier	Accuracy (%)
[34]	No official name	English, Kannada, Tamil, Telugu	S = 200	Adaptive multi-gradient features	Correlation coefficient	87.00
[36]	No official name	English	W = 150	Chi-square	Logistic regression	80.00 76.40
[37]	No official name	Turkish	W = 80	J48 and ID3 algorithms	DT	70.83 93.75

(S is the number of samples, W is the number of writers, and TR, VL, and TS are the number of training, validation, and testing samples, respectively, used for experiments)

accuracy of 94.7% was obtained with local gradient features [27, 28]. It should, however, be noted that geometric and local features are often sensitive to noise, and the resolution of the scanned document images plays a significant role in final prediction results. Moreover, SVM's performance can be affected by the choice of the kernel and its parameters. In addition, although SVM can handle high-dimensional feature spaces effectively, it can be computationally expensive, especially for large datasets.

In addition, open-end-point and curve-fitting based methods were considered for feature extraction in the gender detection literature [20]. Several classification techniques, such as SVM, K-NN, RF, hybridization, and multi-layered perceptron, were further used to detect genders from handwritten documents. The proposed method was evaluated using Gurumukhi (Punjabi) handwriting, and the highest accuracy of 90.57% was obtained with curve fitting-based features and a hybrid classifier [20].

The curvelet transformation and One-class SVMs were further proposed for gender detection [29]. The proposed system was evaluated using the IAM database, and the highest accuracy of 62.49% was obtained. Wavelet transforms using symbolic dynamic filtering were also considered to extract features from each level of document images using a probabilistic finite state automata (PFSA) [21]. The classification was conducted using ANN and SVM classifiers. Various experiments, such as text-dependent, text-independent, script-dependent, and script-independent, using QUWI and MSHD databases, were conducted to evaluate the proposed method. The best gender detection accuracies of 77.70% and 77.60% using script-dependent scenarios with an SVM classifier were obtained from QUWI and MSHD databases, respectively [21]. It is, however, important to note that although wavelets offer a multi-resolution analysis, wavelet transforms are inherently not scale-invariant, and system

performances may be affected by changes in the scale or size of images if the original wavelets are used for handwritten document characterization.

Various features, such as histograms of oriented gradients (HOG), local binary patterns (LBP), and gray-level co-occurrence matrices (GLCM), and ensembles of classification techniques based on SVM, ANN, DT, and RFs were also examined to improve gender classification accuracy in the literature [30]. The authors conducted several experiments on the QUWI database and obtained better results, up to 85% accuracy, using ensemble classifiers. Similarly, in [14], features from handwriting images were extracted using HOG and gradient local binary patterns (GLBP). The SVM classifier was employed for detecting gender from document images. The proposed system was evaluated using the KHATT and IAM databases, and an accuracy of around 70% was reported while the databases were combined for language-independent tests. In [31], numerous local features, such as pixel distribution, pixel density, LBP, and HOG, were considered for the gender prediction task. The prediction was conducted using the SVM classifier. Two different training sets using the IAM database were employed to evaluate and compare the proposed system. The HOG feature extraction with 70% and 74% correct gender prediction performed better in both sets compared to the other methods. The same authors further examined gender, age, and handedness detection tasks using HOG and pixel density features [32]. Considering the IAM database, an accuracy of 73.63% for gender detection and 73.21% for age detection using pixel density was reported. The LBP method was recently extended to AND local binary pattern (ALBP) and OR local binary pattern (OLBP) to extract two-level neighboring pixels from handwriting images in order to improve gender prediction tasks [33]. The highest accuracy was obtained using the

ALBP feature extraction method and the SVM classifier [33].

In [34], Adaptive Multi-Gradient (AMG) features were also extracted using adaptive multi-gradient of Sobel kernels for the gender identification task. The text lines in each image were extracted by considering the dominant pixels. The correlation coefficient between lines and their consistency and inconsistency were computed to identify the gender. Finally, by finding converging and diverging criteria using the error of correlation between the first and successive text lines, the gender of the writer was identified [34]. It is worth mentioning that the performance of AMG may be sensitive to various parameters, such as the geometry of the domain. This method also requires tuning parameters to obtain optimal convergence rates and accuracy.

In addition to other texture features, the Gabor based feature extraction method was used for gender detection in the literature [13]. The mean and standard deviation values of the filtered images were computed and then the Fourier transform was employed to obtain a feature set. The classification was carried out using ANN. The QUWI database was considered for evaluating the proposed system. By employing the same protocol as the ICDAR 2015, the highest accuracy of 70% was reported [13]. Gender detection using textural information obtained based on oriented basic image features (oBIFs) was proposed in [6]. The SVM was applied on three subsets of the QUWI database using experimental protocols of the ICDAR 2013, ICDAR 2015, and ICFHR 2016, and accuracies of 71%, 76% and 68% were obtained, respectively.

The literature on gender detection shows that using only texture feature extraction methods, such as LBP, GLCM, HOG, and GLBP, cannot provide high detection accuracies. The extracted features using these methods are generally converted into histograms which decrease individuality. Moreover, sensitivity to image rotation, noise, resolution, and grayscale are other weaknesses of texture features.

In addition to texture features, the significance of Cloud of Line Distribution (COLD) and Hinge features, along with an SVM classifier, was investigated in the gender detection literature [7]. The proposed system was then evaluated using the QUWI database, and the highest accuracy of 64.40% was reported [7]. Graphological features, such as height, pressure, and margin, followed by a fuzzy rule-based classification, were also proposed for gender detection [35]. A small number of samples (75 digital samples) were considered to evaluate the proposed method, and an accuracy of 76% correct gender detection was reported [35].

In [36], likelihood ratios and binomial logistic regression were considered for gender detection from handwriting. The chi-square test was employed for feature extraction. Applying the proposed method on a dataset written by 150 individuals resulted in a correct classification rate of 80% for females and 76.4% for males [36]. Decision trees and data mining techniques in conjunction with J48 and ID3 algorithms were proposed for identifying gender from handwriting [37]. Employing the decision tree with the J48 and ID3 algorithms on the database, an accuracy of 70.83% and 93.75% were obtained, respectively.

Gender prediction based on static and dynamic features and classification methods, such as KNN, SVM, and Naïve Bayes, was also proposed in the literature [38]. The SVM classifier performed better compared to the other classifiers. Gender detection based on spatial pyramid matching was further proposed in [39]. The weighted histogram of the SIFT descriptor was considered for extracting features from sub-regions. For detecting the gender of a writer from a document image, SVM and ensemble classifiers were then employed. The proposed system was evaluated using QUWI and MSHD databases in script-dependent and script-independent scenarios. Better accuracies of 82% and 90% were obtained using an ensemble classifier in script-dependent scenarios considering QUWI and MSHD databases, respectively. Moreover, features such as pen pressure, margins, irregularity, and space between words and SVM and ANN classifiers, were considered for gender detection [40]. The proposed method was evaluated using the FSHS database, and the gender detection accuracies of 94.7% and 97.1% were obtained using SVM and ANN, respectively [40].

The traditional gender detection approaches based on online methods are comparably fewer than offline methods. An online gender prediction based on an allographic approach was proposed in the literature [41]. Uppercase characters in Spanish were collected using an online device, and both pen-down and pen-up strokes, as the structural features, were considered to characterize online handwriting. The highest accuracy of 74% was reported [41]. In addition, other features, such as time, space, ductus, and pressure, were considered for online gender detection in the literature [10]. A collection of 240 online handwriting and drawing samples using a digital pen and Wacom tablets was created to help with gender, male and female, detection in online handwriting [10]. Furthermore, Marzinotto et al. [15] proposed two feature sets, including dynamic and spatial, for age and gender detection. The dynamic features consist of acceleration, speed, and jerk, while the spatial features consist of local pen trajectories. A two-level clustering approach using the K-means algorithm and the Bag of Prototype Words (BPW) was proposed for age and gender detection [15]. The IRONOFF, an online

database composed of English and French words, was considered to evaluate the proposed method [15].

4.2.2 Deep learning methods

Deep learning generally uses several hidden layers to create a deep neural network for learning features and patterns to make intelligent decisions. Deep learning methods have grown substantially as they provide more accurate results with large databases compared to traditional methods. Data augmentation techniques and transfer learning have further been employed to enable the application of deep learning, where a small number of samples are available for training [42].

Researchers from different disciplines, including the handwriting analysis community, have recently used and explored deep learning in different applications [43, 44]. Table 3 shows the recent research on gender detection using deep learning methods. The results are listed based on the databases used for experimental analysis.

As demonstrated in Table 3, deep learning was used for feature extraction and classification in the gender detection literature. For example, two convolutional neural network-based methods, ResNet and GoogleNet, were applied on Arabic handwritten samples to extract deep features [42]. An SVM was then considered for gender detection purposes. The ResNet obtained the best accuracy of 83.32% on the FSHS database [42]. AlexNet was further used at different levels of documents, including word, patch, and the whole page, to extract deep features for gender detection [45]. For experimental analysis, Linear Discriminant Analysis (LDA) was employed for detection/classification purposes. The QUWI database and different scenarios were considered to evaluate the proposed method, and the highest accuracy of 70.08% correct gender prediction was obtained [45]. In [43], a network architecture consisting of four convolutional layers, a single fully-connected layer, and a softmax output layer was proposed for the final gender classification. The majority vote and average softmax were used for classification purposes. Several experiments were conducted on the Hebrew database considering intra, inter, and mixed-language setups to evaluate the proposed system. The highest accuracy of 82.89% was reported using mixed language with the majority vote classification [43]. Recently, bilinear ResNet (B-ResNet), followed by a softmax layer, was proposed to extract fine-grained features from offline handwriting and perform age and gender detection [46]. To evaluate the

Table 3 Overview of the recent research on gender detection using deep learning methods

Study	Database	Script	Setup	Feature	Classifier	Accuracy %
[45]	QUWI	English Arabic	TR = 2800, TS = 1200	AlexNet	LDA	70.08
[47]	IAM KHATT	English Arabic	TR = 8462, TS = 2464 TR = 5777, TS = 967	Rectification units (ReLU), softmax	CNN	80.72 68.90
[48]	IAM KHATT	English Arabic	S = 330 S = 270	InceptionV3, DenseNet201, Xception	CNN	84.00 75.00
[44]	IAM KHATT ICDAR 2013	English Arabic	TR = 2573, TS = 1286 TR = 1333, TS = 667 TR = 1128, TS = 772	ATP-DenseNet-169	CNN	77.60 74.10 71.80
[46]	KHATT HHD	English Arabic Hebrew	TR = 1400, VL = 300, TS = 300 TR = 560, VL = 72, TS = 70	B-ResNet	B-CNN	76.17 84
[43]	HEBIU	Hebrew-English	TR = 268, VL = 38, TS = 76	CNN	Majority vote, Average softmax	82.89
[42]	FSHS	Arabic	S = 2700	ResNet	SVM	83.35

(S is the number of samples, TR, VL, and TS are the number of training, validation, and testing used in the experiments)

proposed system, experiments were conducted on the KHATT and HHD (Hebrew) databases, and accuracies of 76.17% and 84% were obtained, respectively [46]. In addition, ATP-DenseNet, with two pathways, feature pyramid, and A-DenseNet, was developed for the problem of gender detection [44]. Page-level features were extracted using a feature pyramid, while the word-level features were extracted using A-DenseNet. IAM, KHATT, and ICDAR 2013 databases were considered to evaluate the proposed method, and accuracies of 77.6%, 74.1%, and 71.8% were obtained, respectively [44].

In some research work, handwriting samples were used for classifying more than one biometric characteristic, such as gender and handedness, of a writer [47]. A CNN architecture was considered for feature extraction and classification processes. The IAM and KHATT databases were used for experiments, and accuracies of 80.72% and 68.90% were reported, respectively [3]. Moreover, deep learning methods, including InceptionV3, DenseNet201, and Xception, were used for gender and handedness classification [48]. The IAM and KHATT databases were considered for experimental analysis, and the DenseNet201 performed better than other models. It is worth mentioning that the availability of enough training data is one of the key parameters to further improve the accuracies of deep learning based age/gender/handedness detection approaches. Therefore, using robots for generating human-like data may be one area of research focus in this niche research field.

4.3 Word frequency graph of the related works

We performed the quantitative analysis to determine the trend of handwriting analysis between 2012 and 2023 to find the methods and databases used frequently in the literature. The KH Coder software (https://khcoder.net/en/) helps explore and analyze large amounts of unstructured and semi-structured text to distinguish concepts, topics, patterns, and other aspects of interest. The KH Coder can perform word frequency statistics, automatic clustering, and analyze survey data, such as questionnaires and interviews, based on text mining analysis [49, 50].

This study used the KH Coder to analyze metadata collected through our research [49]. The general steps in data mining using KH Coders are collecting and organizing the text, loading the text into software, filtering the words using chi-square value, and then generating the co-occurring relationship. We used the KH Coder to conduct a textual analysis of the title, abstract, and keywords of 42 selected papers related to age and gender detection problems.

A co-occurrence network diagram is drawn based on retrieved words with similar appearance patterns, where nodes/circles show the retrieved words and the lines (edges) show the interconnectivity between words. In the co-occurrence network diagram, the words with high degrees of co-occurrence are connected through several lines. The size of the circles shows the frequency of occurrence of the words, so larger circles appear more frequently than smaller circles. Figure 4 crafts an appropriate model with 14 subgraphs or clusters to describe relationships between topics. For example, the cluster with higher word frequency (i.e., colored in orange) is composed of related words such as "handwriting", "gender", "classification", "feature", and "age". The interconnection between the words, such as "age-category", "feature-classification", and "gender-feature-classification," informed the derivation of themes. These visualizations demonstrate the results according to the keywords of the papers. It should also be noted that the frequency of the word "gender" is higher than the word "age," which means more research on gender detection was carried out in the last decade. Similarly, the commonly used databases are the QUWI and IAM databases. In addition, English and Arabic scripts are among the most frequently used scripts for age and gender detection in the literature.

Another co-occurrence network based on the year of publication is shown in Fig. 5. Figure 5 demonstrates that the words "handwriting," "gender," "classification," and "feature" were frequently used between 2012 and 2023. In addition, offline methods were more often explored in 2012, whereas online methods were increasingly explored in 2020. Texture features and SVM-based classifiers were investigated highly in 2017, whereas convolutional neural networks received more attention since 2018. Age and handedness classification received some attention in 2015.

5 Metrics and databases

Databases and evaluation metrics are pillars to fairly compare state-of-the-art methods' performances in any research domain. Accuracy and precision, as two evaluation metrics, have been commonly used in the literature for age and gender classification/detection tasks. Accuracy is the number of correct classified/detected samples divided by the total number of samples. Precision is the fraction of the relevant samples among the retrieved samples.

The process of developing standard handwritten databases was initially commenced when research in handwriting/document analysis and recognition started in the early 1990s and received significant attention in the last three decades [51]. The databases have been commonly created by requiring the subjects and writers to fill in a questionnaire with their personal information (metadata), such as gender, age category, handedness, nationality,
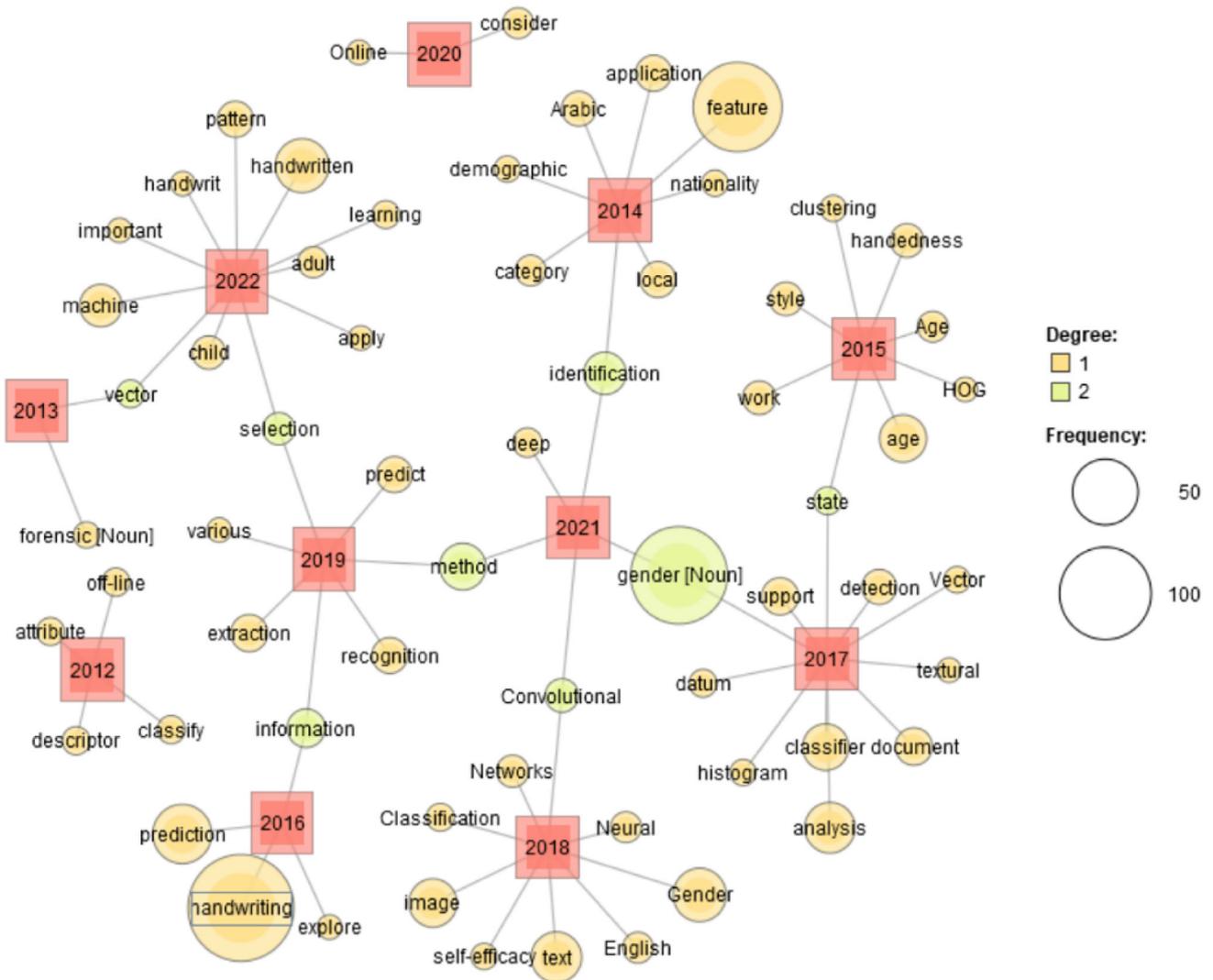
Fig. 4 The co-occurrence network words diagram

education level, and profession. There are a few handwriting databases with a different number of writers and samples written in a single script or multiple scripts for gender, age, and handedness in the literature [52–55].

Most publicly accessible databases are offline databases. In offline databases, writers copied a text (multiple texts) on separate sheets of paper using pens. Then, documents were scanned using scanners or cameras and converted into image formats, such as PNG, JPEG, GIF, or TIFF. A few online available databases were also collected using the Wacom tablets. The lower availability of online databases compared to offline databases may refer to the cost of required devices for collecting online handwriting. In online databases, in addition to handwriting, other features, including pen pressure, angle, speed, and time in the air, were collected to be used for further analysis. A list of databases used in the literature for age and gender detection is provided in Table 4. Table 4 reveals that a relatively

larger number of samples were collected in offline databases than in online databases. Further analyses show that most of the available databases were written in English, Arabic, and French scripts.

IAM, QUWI, KHATT, MSHD, and ICDAR are a few publicly accessible databases commonly used for offline age and gender detection in the literature [27, 29, 30]. The IAM handwriting database includes 1539 English handwritten document pages written by 657 writers [55]. Documents were scanned at a resolution of 300 dpi in PNG format. The lines and words on each page were also segmented and made available in the same format for detailed analysis [55]. The Qatar University Writer Identification (QUWI) database [53] contains Arabic and English handwritten documents. This database is a collection of 1,017 handwriting samples. Each writer contributed two Arabic and two English texts, including two arbitrary and two predefined samples. Documents were scanned at a

Fig. 5 The co-occurrence network diagram obtained based on the year of publications

resolution of 600 dpi in JPEG format [53]. The KHATT is an offline Arabic database [54] of handwritten texts written by 1000 writers in different age categories. Each writer completed a form of four pages, and documents were scanned at resolutions of 200, 300, and 600 dpi. The written samples include 2000 randomly chosen paragraphs from 46 sources covering all Arabic characters [54]. The multi-script handwritten database (MSHD) [52] contains 1300 Arabic and French handwriting samples written by 100 writers. Each writer copied six texts in Arabic, six in French, and one page of digits which can be used for age and gender analysis [52].

It is noted that MSHD and KHATT databases are suitable for text-dependent, text-independent, script-dependent, and script-independent experiments. In script-dependent experiments, training and testing samples are from the same scripts; however, in script-independent

experiments, training and testing samples are from different scripts. For instance, a system can be trained with Arabic document images and tested with French/English document images and vice versa.

In addition to the above datasets, several competitions were organized over the last decade in conjunction with ICDAR 2009, ICDAR 2011, ICDAR 2013, and ICDAR 2015 for writer, gender, and age detection. Most databases used in these competitions are publicly available to extend research in this field. However, it is worth mentioning that due to the current research trend in deep learning methods, the number of samples in all benchmark databases is relatively small and insufficient for effectively training different models, including deep learning models. Thus, creating databases with a large number of writers and samples would be beneficial for designing more complex

Table 4 Handwriting databases used for age and gender detection in the literature

Type	Database (URL to access)	Writer	No. Samples	Script
Offline	IAM [55]	657	1539	English
	(https://fki.tic.heia-fr.ch/databases)			
	QUWI [53]	1017	4068	English—Arabic
	(via email)			
	KHATT [54]	1000	4000	Arabic
	(http://khatt.ideas2serve.net/KHATTDownload.php)			
	MSHD [52]	100	1300	French—Arabic
	(https://www.kaggle.com/datasets/chawkidjeddi/lamismshd-offline-handwriting-database)			
	FSHS [56]	2200	2700	Arabic
	(via email)			
	HEBIU [43]	405	810	Hebrew—English
	(Private)			
	ICDAR2015	1017	4068	English—Arabic
	(https://paperswithcode.com/dataset/icdar-2015)			
Online	IRONOFF [9]	793		English—French
	(via email)			
	BiosecurID [57]	400		Spanish
	(http://atvs.ii.uam.es/atvs/biosecurid_sonof_db.html)			

deep learning-based methods to obtain more accurate results in this domain.

6 Discussion and remarks

In terms of benchmark and publicly available databases for gender/age detection, three commonly used benchmark databases (IAM, QUWI, and KHATT) are available in the literature. Apart from their limitation in sample size, these databases include only Arabic, English, and French handwritten document image, indicating a lack of databases for other scripts and the unavailability of large-size databases in the literature. Thus, further research on benchmark and evaluation metrics is necessary in this domain.

For detection and classification purposes, artificial neural networks, convolutional deep neural networks, DT, RF, and SVMs were frequently used in the age and gender detection literature (Tables 2 and 3). Although a direct comparison of the results obtained from various methods on different databases is impossible, results reported in the literature have further brought to our attention a noticeable difference between the results of traditional and deep learning methods in the gender classification task. The highest accuracy of 97.10% was obtained among the traditional methods in the literature using pen pressure, margins, irregularity, and space between words features and an ANN classifier on the FSHS database [40].

Moreover, the authors applied their proposed method on ICDAR2013, and an accuracy of 91.4% was reported with the SVM classifier [40]. The lowest accuracy is 66.3% using geometrical features and SVM for the classification, where the ICDAR 2013 database was used for evaluation [23]. Among the deep learning methods, the highest reported accuracy is 84% using the DenseNet201method applied on the IAM database [48]. The lowest accuracy was reported as 68.90% using ReLU and softmax methods applied on KHATT databases [47]. In both studies, CNN was used for the classification. The literature review indicates that in most studies, traditional models, especially SVM-based methods, performed better than the other models, including deep learning-based approaches [40]. The lower accuracy obtained using deep learning compared to SVMs and other traditional models could be due to the following reasons: limitations in the number of samples and/or the lack of diversity in databases to train high-performance models using deep learning approaches to learn meaningful patterns, and noise sensitivity of deep learning models, as they are highly data-driven and can be sensitive to noisy or outlier data points and if a dataset contains significant amounts of noise, deep learning models may be unable to make reliable predictions.

From another perspective, we reviewed Tables 2 and 3 and compared the results obtained based on the databases. It is important to note that the highest accuracy on QUWI was 85% when texture features and ensemble classifiers

were used for gender detection [30]. Meanwhile, the highest accuracy of 90% was obtained on the MSHD when SIFT descriptors, spatial pyramid matching, and ensemble classifiers were used for gender detection [39]. The highest accuracy of 94.79% on the ICDAR 2013 database was obtained when Gradient and WD-LBP features and SVMs were used for gender detection [27]. The highest gender detection accuracy of 97.1% was obtained on the FSHS using pen pressure, margins, and irregularity features [40]. Considering the IAM dataset, the highest accuracy of 84% was obtained based on InceptionV3, DenseNet201, and Xception feature extraction methods and CNN classifiers [48]. The highest accuracy of 76.17% was obtained on the KHATT database when B-ResNet was considered for feature extraction, and B-CNNs were employed for the classification [46]. It is worth mentioning that the traditional methods performed relatively better on QUWI, MSHD, FSHS, and ICDAR 2013 databases, whereas deep learning performed better on IAM and KHATT databases.

At the outset, it is important to note that deep learning achieved remarkable success in many research domains and applications. However, it is not a one-size-fits-all solution, and depending on the specific problem, dataset, resources, interpretability requirements, or domain expertise, traditional methods may still be preferred or even outperform deep learning approaches. Thus, there would be many opportunities for researchers in this field to expand the investigation and research on age and gender classification/detection problems using deep learning methods to obtain more accurate results.

7 Conclusion

This article presented a systematic and comprehensive review of age and gender classification/detection from handwritten documents. A general age and gender identification framework comprising pre-processing, feature extraction, and classification was provided and discussed in detail. The most commonly used databases, their number of samples, and the type of scrips were discussed. The methods in the literature were grouped into traditional and deep learning methods, and the results were compared accordingly. The KH Coder software was used for text mining purposes, and the trends of words in the literature were analyzed.

This study can bring the requirements of handwriting analyses to the attention of novice researchers and open a new direction for them. Although there has been considerable research on this area (age and gender detection from handwriting), it is still considered a challenging problem. Neither computerized analysis nor human experts achieved highly accurate results for these tasks. In further research

on these problems, researchers can investigate the suitabilities of different kinds of features to characterize gender and age from handwriting and also explore feature selection techniques to identify the most appropriate features for these problems. These types of features would be beneficial, as the annotated datasets in these areas are still of small size. Deep learning, especially transfer learning and continuous adaptive learning, is another direction of research to improve age/gender detection accuracies. Synthetic data generation, data augmentation, and a combination of these methods may also be investigated by researchers in the future to explore different results and build plausible systems.

Declarations

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Chaubey G, Arjaria SK (2022) Personality prediction through handwriting analysis using convolutional neural networks, In: Proceedings of the International Conference on Computational Intelligence, pp 59–70
2. Dargan S, Kumar M (2020) A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities. Expert Syst Appl 143:113114
3. Alaei F, Alaei A (2022) Handwriting analysis: applications in person identification and forensic Breakthroughs in digital biometrics and forensics. Springer, Cham, pp 147–165
4. Beech JR, Mackintosh IC (2005) Do differences in sex hormones affect handwriting style? Evidence from digit ratio and sex role identity as determinants of the sex of handwriting. Personal Individ Differ 39(2):459–468

5. Burr V (2002) Judging gender from samples of adult handwriting: accuracy and use of cues. J Soc Psychol 142(6):691–700

6. Gattal A, Djeddi C, Siddiqi I, Chibani Y (2018) Gender classification from offline multi-script handwriting images using oriented Basic Image Features (oBIFs). Expert Syst Appl 99:155–167. https://doi.org/10.1016/j.eswa.2018.01.038

7. Gattal A, Djeddi C, Bensefia A, Ennaji A (2020) Handwriting based gender classification using COLD and hinge features, in Image and Signal Processing. Springer, Cham, pp 233–242

8. Marzinotto G et al. (2016) Age-related evolution patterns in online handwriting, in Computational and mathematical methods in medicine, vol 2016

9. Viard-Gaudin C, Lallican PM, Knerr S, and Binter P (1999) The IRESTE On/Off (IRONOFF) dual handwriting database, In: Proceedings of the International Conference on Document Analysis and Recognition. pp 455–458. doi: https://doi.org/10.1109/ICDAR.1999.791823.

10. Cordasco G, Buonanno M, Faundez-Zanuy M, Riviello MT, Likforman-Sulem L, and Esposito A (2020) Gender identification through handwriting: an online approach, In: Proceedings of the International Conference on Cognitive Infocommunications, pp 000197–000202

11. Srihari SN, Huang C, Srinivasan H, Shah V (2007) Biometric and forensic aspects of digital document processing in Digital Document Processing. Springer, London, pp 379–405

12. Basavaraja V, Shivakumara P, Guru DS, Pal U, Lu T, and Blumenstein M (2019) Age estimation using disconnectedness features in handwriting, In: Proceedings of the International Conference on Document Analysis and Recognition pp 1131–1136. doi: https://doi.org/10.1109/ICDAR.2019.00183.

13. Mirza A, Moetesum M, Siddiqi I, and Djeddi C, (2016) Gender classification from offline handwriting images using textural features, In: Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp 395–398. doi: https://doi.org/10.1109/ICFHR.2016.0080

14. Bouadjenek N, Nemmour H, and Chibani Y (2015) Age, gender and handedness prediction from handwriting using gradient features, In: Proceedings of the International Conference on Document Analysis and Recognition, pp 1116–1120. doi: https://doi.org/10.1109/ICDAR.2015.7333934.

15. Marzinotto G, Rosales JC, El-Yacoubi MA, and Garcia-Salicetti S (2015) Age and gender characterization through a two-layer clustering of online handwriting, In Proceedings of the International Conference on Advanced Concepts for Intelligent Vision Systems, pp 428–439

16. AL-Qawasmeh N, Suen CY (2022) Transfer learning to detect age from handwriting, Advances in Artificial Intelligence and Machine. Learning 2:394–406

17. Shin J et al (2022) Important features selection and classification of adult and child from handwriting using machine learning methods. Appl Sci 12(10):10

18. Rosales JC, Marzinotto G, El-Yacoubi MA, and Garcia-Salicetti, S (2015) Age characterization from online handwriting, In: International Symposium on Pervasive Computing Paradigms for Mental Health, pp176–185

19. AL-Qawasmeh N, Khayyat M, Suen CY (2023) Age detection from handwriting using different feature classification models. Pattern Recognit Lett 167:60–68

20. Dargan S and Kumar M (2021) Gender classification and writer identification system based on handwriting in Gurumukhi script, In: Proceedings of the International Conference on Computing, Communication, and Intelligent Systems, pp 388–393, https://r12a.github.io/scripts/gurmukhi

21. Akbari Y, Nouri K, Sadri J, Djeddi C, Siddiqi I (2017) Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. Image Vis Comput 59:17–30. https://doi.org/10.1016/j.imavis.2016.11.017

22. Bi N, Suen CY, Nobile N, Tan J (2019) A multi-feature selection approach for gender identification of handwriting based on kernel mutual information. Pattern Recogn Lett 121:123–132. https://doi.org/10.1016/j.patrec.2018.05.005

23. Tan J, Bi N, Suen CY, and Nobile N (2016) Multi-feature selection of handwriting for gender identification using mutual information, In Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp 578–583

24. Sokic E, Salihbegovic A, and Ahic-Djokic M (2012) Analysis of off-line handwritten text samples of different gender using shape descriptors, In Proceedings of the International Symposium on Telecommunications, pp 1–6

25. Al Maadeed S, Hassaine A (2014) Automatic prediction of age, gender, and nationality in offline handwriting. EURASIP J Image Video Process 2014(1):1. https://doi.org/10.1186/1687-5281-2014-10

26. Siddiqi I, Djeddi C, Raza A, Souici-mesylate L (2015) Automatic analysis of handwriting for gender classification. Pattern Anal Applications 18(4):4. https://doi.org/10.1007/s10044-014-0371-0

27. Ibrahim AS, Youssef AE, and Abbott AL (2014) Global vs. local features for gender identification using Arabic and English handwriting, In: International Symposium on Signal Processing and Information Technology, pp 000155–000160

28. Youssef AE, Ibrahim AS, and Abbott AL (2013) Automated gender identification for Arabic and English handwriting

29. Guerbai Y, Chibani Y, and Hadjadji B (2017) Handwriting gender recognition system based on the one-class support vector machines, In: Proceedings of the International Conference on Image Processing Theory, Tools and Applications, pp 1–5

30. Ahmed M, Rasool AG, Afzal H, Siddiqi I (2017) Improving handwriting based gender classification using ensemble classifiers. Expert Syst Appl 85:158–168. https://doi.org/10.1016/j.eswa.2017.05.033

31. Bouadjenek N, Nemmour H, and Chibani Y (2014) Local descriptors to improve off-line handwriting-based gender prediction, In: Proceedings of the International Conference of Soft Computing and Pattern Recognition pp 43–47. doi: https://doi.org/10.1109/SOCPAR.2014.7007979

32. Bouadjenek N, Nemmour H, and Chibani Y (2015) Histogram of oriented gradients for writer's gender, handedness and age prediction, In: Proceedings of the International Symposium on Innovations in Intelligent Systems and Applications, pp 1–5

33. Verma A, Rahi R, Singh NP (2023) Novel ALBP and OLBP features for gender prediction from offline handwriting. Int J Inf Technol 15(3):1453–1464

34. Navya BJ et al. (2018) Adaptive multi-gradient kernels for handwriting based gender identification, In: Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp 392–397. doi: https://doi.org/10.1109/ICFHR-2018.2018.00075

35. Riza LS, Zainafif A, Rasim R, Nazir S (2018) Fuzzy rule-based classification systems for the gender prediction from handwriting. Telecommun Comput Electron Control 16(6):6

36. Sharma V et al (2021) Novel use of logistic regression and likelihood ratios for the estimation of gender of the writer from a database of handwriting features. Aust J Forensic Sci. https://doi.org/10.1080/00450618.2021.1956587

37. Topaloglu M, Ekmekci S (2017) Gender detection and identifying one's handwriting with handwriting analysis. Expert Syst Appl 79:236–243. https://doi.org/10.1016/j.eswa.2017.03.001

38. Erbilek M, Fairhurst M, and Li C (2016) Exploring gender prediction from digital handwriting, In: Proceedings of the Signal Processing and Communication Application Conference, pp 789–792

39. Alaei F and Alaei A (2021) Gender detection based on spatial pyramid matching, In: Proceedings of the International Conference on Document Analysis and Recognition, pp 305–317

40. Alqawasmeh N, Khayyat M, Suen CY (2022) Novel features to detect gender from handwritten documents. Pattern Recognit Lett. https://doi.org/10.1016/j.patrec.2022.08.016

41. Sesa-Nogueras E, Faundez-Zanuy M, Roure-Alcobé J (2016) Gender classification by means of online uppercase handwriting: a text-dependent allographic approach. Cognit Comput 8(1):1

42. AL-Qawasmeh N and Suen CY (2022) Gender detection from handwritten documents using transfer learning method, In: Advances in Pattern Recognition and Artificial Intelligence. World Scientific, doi:https://doi.org/10.1142/9789811239014_0007

43. Illouz E, David E, Netanyahu NS (2018) Handwriting based gender classification using end-to-end deep neural networks in Artificial Neural Networks and Machine Learning. Springer, Cham, pp 613–621

44. Xue G, Liu S, Gong D, Ma Y (2021) ATP-DenseNet: a hybrid deep learning-based gender identification of handwriting. Neural Comput Appl 33(10):4611–4622

45. Moetesum M, Siddiqi I, Djeddi C, Hannad Y, and Al-Maadeed S (2018) Data driven feature extraction for gender classification using multi-script handwritten texts, In: Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp 564–569

46. Rabaev I, Alkoran I, Wattad O, Litvak M (2022) Automatic gender and age classification from offline handwriting with bilinear ResNet. Sensor 22(24):1–12. https://doi.org/10.3390/s22249650

47. Morera Á, Sánchez Á, Vélez JF, and Moreno AB (2018) Gender and handedness prediction from offline handwriting using convolutional neural networks, In: Complexity, vol. 2018

48. Rahmanian M, Shayegan MA (2021) Handwriting-based gender and handedness classification using convolutional neural networks. Multimed Tools Appl 80(28):28

49. Higuchi K (2016) KH Coder 3 reference manual. Ritsumeikan University, Kioto (Japan)

50. Aboagye-Otchere F, Agyenim-Boateng C, Enusah A, Aryee TE (2021) A review of big data research in accounting, intelligent systems in accounting. Financ Manag 28(4):268–283. https://doi.org/10.1002/isaf.1504

51. Hussain R, Raza A, Siddiqi I, Khurshid K, Djeddi C (2015) A comprehensive survey of handwritten document benchmarks: structure, usage and evaluation. EURASIP J Image Video Process 2015(1):1–24

52. Djeddi C, Gattal A, Souici-Meslati L, Siddiqi I, Chibani Y, and Abed HE (2014) LAMIS-MSHD: a multi-script offline handwriting database, In: Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp 93–97. doi: https://doi.org/10.1109/ICFHR.2014.23

53. Maadeed SA, Ayouby W, Hassaïne A, and Aljaam JM (2012) QUWI: An Arabic and English handwriting dataset for offline writer identification, In: Proceedings of the International Conference on Frontiers in Handwriting Recognition, pp 746–751. doi: https://doi.org/10.1109/ICFHR.2012.256

54. Mahmoud SA et al (2014) KHATT: an open Arabic offline handwritten text database. Pattern Recognit 47(3):3

55. Marti U-V, Bunke H (2002) The IAM-database: an English sentence database for offline handwriting recognition. Int J Doc Anal Recognit 5(1):1

56. AL-Qawasmeh N, and Suen CY (2020) Gender detection from handwritten documents using concept of transfer-learning, In: Proceedings of the International Conference on Pattern Recognition and Artificial Intelligence, pp 3–13

57. Fierrez J et al (2010) BiosecurID: a multimodal biometric database. Pattern Anal Appl 13(2):235–246