DOI: 10.17986/blm.1690
Adli Tıp Bülteni 2024;29(1):9-19

Sex Determination Using Data Mining Methods Through Measurements of Ascender and Descender Parts of Letters

Harflerin Alçalan ve Yükselen Uzantılarının Ölçümleri Kullanılarak Cinsiyetin Veri Madenciliği Yöntemleriyle Belirlenmesi

Dilara Öner Kaya[1], Yasin Koca[2], Tuğba Ülker Kuzubaş[3], Ömer Kurtaş[4], İbrahim Demir[5], Gürsel Çetin[6]

[1]Independent Researcher, Istanbul, Turkey
[2]Council of Forensic Medicine, Istanbul, Turkey
[3]Adana City Training and Research Hospital, Forensic Medicine Polyclinic, Adana, Turkey
[4]Kocaeli University Faculty of Medicine, Department of Forensic Medicine, Kocaeli, Turkey
[5]Turkish Statistical Institute, Ankara, Turkey
[6]Istanbul University-Cerrahpasa, Cerrahpasa Faculty of Medicine, Department of Forensic Medicine, Istanbul, Turkey

ABSTRACT

Objective: Sex determination has been found interesting in forensic handwriting examinations and has been researched by scientists. The inclusion of the sex parameter as a supporting element in the examination of forensic handwriting while deciding belonging will increase the reliability of the results. In addition, it will help reduce the number of people to be investigated for a large group of suspects, both men and women. In this study, it was aimed to investigate the contribution of the ascender and descender parts of the letters to sex prediction by measuring them.

Methods: In line with this purpose, handwriting samples were collected from 50 female and 50 male participants by having them write 11 sentences containing the letters "b, d, f, g, h, k, t, y, p" at initial, medial, and end positions. The ascender and descender parts of these letters were measured in millimeters. Logistics, k-nearest neighbor (KNN), support vector machine (SVM) and artificial neural network (ANN) were selected and applied to these data.

Results: The ascender and descender parts of these letters were measured in millimeters and statistically significant differences were found between male and female participants. The ascender parts of the "b, d, h, k, t" were determined to be statistically significantly longer in males. Accuracy rates are 0.65, 0.60, 0.71 and 0.82 for Logistics, KNN, SVM and ANN, respectively.

Conclusion: In our opinion, this result is promising. If the studies on this subject are increased, higher success rates can be achieved, and more contributions can be made to forensic handwriting examination.

Keywords: Sex prediction, forensic handwriting examination, artificial neural networks, data mining

***A part of this study was presented as an oral presentation with the title "Gender Estimation in Forensic Handwriting Examinations with Artificial Neural Networks Analysis" at the 3rd Forensic Document Examination Congress, held online on 05-06 November 2021.**

Address for Correspondence/Yazışma Adresi: Dilara Öner Kaya, Independent Researcher, Istanbul, Turkey
E-mail: drdilaraoner@gmail.com
ORCID ID: orcid.org/0000-0001-7478-3720

Received/Geliş tarihi: 29.12.2023
Accepted/Kabul tarihi: 06.02.2024

ÖZ

Amaç: Adli el yazısı incelemelerinde cinsiyet tahmini yıllar boyunca ilgi çekici bulunmuş ve bilim insanları tarafından araştırılmıştır. Aidiyet kararı verilirken adli el yazısı incelemesinde diğer tanı unsurlarının yanında cinsiyet parametresinin destekleyici bir unsur olarak yer alması sonuçların güvenilirliğini artıracaktır. Ayrıca hem erkek hem de kadınlardan oluşan geniş bir şüpheli grubu için araştırılacak kişi sayısının azaltılmasına da yardımcı olacaktır. Bu nedenle adli el yazısı incelemesinde cinsiyet tahmini üzerine araştırmalar yapılmaktadır. Bu çalışmada belirlenen harflerin alçalan ve yükselen kısımlarının ölçülerek cinsiyet tahminine katkısının araştırılması amaçlanmıştır.

Yöntem: Bu amaç doğrultusunda Yükseköğretim öğrencisi ve mezunu 50 kadın ve 50 erkek katılımcıdan, Türk dilinde başta ortada ve sonda olacak şekilde "b, d, f, g, h, k, t, y, p" harflerini içeren 11 adet cümle yazdırılmak sureti ile yazı örnekleri toplanmıştır. Bu harflerin yükselen ve alçalan kısımları milimetre cinsinden ölçülmüştür. Makine öğrenmesi çalışmalarında en başarılı dört yöntem olan logistik regresyon, K en yakın komşuluk (KNN), destek vektör makineleri (DVM) ve yapay sinir ağları (YSA) seçilerek bu verilere uygulanmış ve yazıyı yazan kişinin yazı karakterine göre erkek ve kadın olup olmadığı belirlenmeye çalışılmıştır.

Bulgular: Harflerin alçalan ve yükselen kısımlarının ölçülmesi ile kadın ve erkek katılımcılar arasında istatistiksel olarak anlamlı farklılıklar bulunmuştur. Erkeklerde "b, d, h, k, t'nin" yükselen kısımlarının istatistiksel olarak anlamlı derecede daha uzun olduğu belirlenmiştir. Doğruluk oranları lojistik, KNN, DVM ve YSA için sırasıyla 0,65, 0,60, 0,71 ve 0,82'dir.

Sonuç: Kanaatimizce elde edilen sonuçlar ümit vericidir. Bu konudaki çalışmaların artırılması halinde daha yüksek başarı oranları elde edilebilir ve adli el yazısı incelemesine daha fazla katkı sağlanabilir.

Anahtar Kelimeler: Cinsiyet tahmini, adli el yazı incelemeleri, yapay sinir ağları, veri madenciliği

INTRODUCTION

Handwriting is affected by many factors such as age, social habits, and biological factors, which is the starting point of studies aimed at classifying handwriting into demographic classes (1). Characteristic elements found in people's handwriting have been investigated for sex determination (2).

In forensic handwriting examinations, belonging is determined by comparing the distinguishing features of handwriting. Huber and Headrick (2) reported that handwriting has 21 distinctive features (3).

In forensic handwriting examinations, the ability to classify the handwriting into demographic data such as age, sex, hand dominance, and nationality and then performing eliminations is of great benefit in practice (1). This classification can help forensic document examiners to focus on a particular category of suspects (4). Further, the classification of demographic information is regarded to be objective because it can be experimentally verified through quantitative results (5). Moreover, by processing all these demographic data separately, improved results can be produced for the determination and verification of the person who wrote the examined handwriting (4).

Although it is a two-class problem, sex determination through handwriting is difficult owing to a large number of variations. The fact that some men have feminine handwriting and vice versa cause significant differences in studies in this area (5).

Determination of certain characteristics through handwriting and the studies that classify them into sex information have generally attracted the interest of psychologists.

The oldest study on this issue was conducted by Goodenough in which 10 female and 10 male graduate students classified the handwritings of 115 high-school students into sex, and approximately 2/3 of the writings were correctly classified into sex (6).

The classification of handwriting into demographic data is conducted in two steps: Feature extraction and classification. The performance of the system has been reported to depend on the feature extraction step because extracted features are used to distinguish individuals (4).

In a study conducted by Marzinotto et al. (7) with two-layer clustering analysis on the writings created online, sex determination was not as good as in the classification of demographic information such as age; the vertical and connected text was observed together in males' writings whereas in females' writings only one or the other form was observed.

In the study conducted by Hamid and Loewenthal (8), handwriting samples were collected from 30 subjects (16 females and 14 males) in both English and Urdu languages, and 25 examiners were asked to classify them into sex; the accuracy was found to be approximately 68%. Multiple analysis of variance was used in this study, and it was reported that language is not an important source of variance in terms of sex. In other words, approximately the same results were obtained in both languages.

In a study conducted by Binet (9), writing samples belonging to a total of 180 participants, 91 male and 89 female, 2 graphologists, and 15 people who were disinterested in such business were asked to classify them into sex. Results indicated that one of the experts classified 78.3% correctly, and 10 disinterested examiners classified correctly at a percentage ranging between 65.9% and 72% (10).

Many features such as the size of the letters, formation features, line, and word spacing, dotting of the letter "i", inclination, and slant of the writings were determined by Kumar et al. (3) using the Z test. They investigated whether a statistical difference existed between male and female writings and reported that a significant difference existed in terms of sex.

Referring to the study conducted by Briggs with 100 people, Tomai et al. (11) reported that the former had concluded that distinguishing male and female writings was not possible. However, one of the frequently studied topics regarding demographic features in the literature is sex determination, and these studies have concluded that a strong relationship exists between sex and certain handwriting features (6,8).

In the literature, there are studies aimed at determining certain features that are thought to be important in sex classification and investigating their relation to sex (4,10). Moreover, studies on automatically classifying sex through artificial neural networks (ANNs) and various image processing techniques (1,5,11-16) using software systems exist.

In the literature research, it was determined that the classification rates of the studies on sex prediction were in the range of 61.93%-85.7% (1,4,5,8,13,15-24). Some of these studies are given in Table 1.

Table 1. Overview of the some studies on gender prediction

Forecasting methodologies	Authors	Result
*ANN	Bandi et al. (1)	Wherein 11 macro features including diagnostic elements such as slant and word spacing were taken into consideration, that a success rate of 73.2% was achieved in gender determination using a single ANN. They also reported that this ratio was further increased to 77.5% by applying ten neural networks and implementing bagging and boosting processes
Measuring different properties of letters and applying Z test	Kumar et al. (3)	Many features such as the size of the letters, formation features, line, and word spacing, dotting of the letter "i," inclination, and slant of the writings were determined by Kumar et al. using the Z test
Rassal random forest and kernel discriminant analysis	Al Maadeed and Hassaine (4)	Reported a classification rate of 73.59% for gender determination using random forest and kernel discriminant analysis, and this rate reached 74.05% when text with identical content was used
Wavelet (Pattern recognition)	Akbari et al. (5)	Success rate of 74% in Arabic and 68% in English and French in a classification of offline texts using a method that treats texts as patterns and uses wavelets to characterize them
MANOVA	Hamid ve Loewenthal (8)	Handwriting samples were collected from 30 subjects (16 females and 14 males) in both English and Urdu languages, and 25 examiners were asked to classify them into gender; the accuracy was found to be approximately 68%.
Non data mining and statistical method	Young (10)	25 female and 25 male untrained judges were asked to classify the writings of 25 male and 25 female participants into gender, and the rate of correct classification ranged from 42% and 72%, with an average of 61%. Results were reported to be 11% better given that the chance factor was 50%
*SVM ANN	Siddiqi et al. (13)	Using support vector machines and ANNs, 68.75% and 73.02% success rates, respectively, were achieved with two different datasets, i.e., the Qatar University Writer Identification and Multiscript Handwritten Database
*SVM	Liwicki et al. (15)	Applying to support vector machine systems and using an online system, the success rates varied between 61.93% and 62.19%
*ANN with performans geliştirici	Ahmed et al. (16)	Many features and classification methods were applied one by one and in combinations: in (a) single comparison, the best performing classifier in English writings was ANNs, and successful classification rates of up to 79% were obtained with performance enhancers
Deep Learning and Convolutional Neural Network	Morera et al. (17)	Deep learning and convolutional neural network were used with Arabic (KHATT) and English (IAM) datasets. They found that the correct classification rate was 68.90% for Arabic and 80.72% for English
*SVM	Youssef et al. (18)	Using support vector machines, a correct classification rate of 68.6% was achieved for Arabic and 85.7% for English when the system was trained separately for English and Arabic. In the system trained for handwriting in both languages, this rate was found to be 74.3%. Moreover, as a conclusion to these results, it was suggested that those analyses could be used independently of language in forensic sciences
Fuzzy Rule-Based Classification	Riza et al. (19)	Using fuzzy rule-based classification systems, a correct classification rate of 76% was achieved considering features such as the pressure level, line, and character height, maximum height in the sentence, baseline, and margins
*ANN	Cha and Srihari (20)	Using ANNs, and the success rate was reported to be 70.2%

Table 1. Continued

Forecasting methodologies	Authors	Result
*SVM	Bouadjenek et al. (23)	Reported that accuracy of approximately 70% was obtained regardless of the language in the gender determination study conducted on both Arabic (KHATT) and English (IAM) handwritten text clusters using histogram oriented gradient and gradient local binary pattern systems and the support vector machine (SVM) classifier
One Class *SVM	Guerbai et al. (24)	Using one-class SVM (OC-SVM), the success rate with a single classifier was reported to be 62.49%, and it was 77.3% in the case of combined classifiers
*ANN SVM LR k-NN	This study	ANN; 82% SVM; 71% LR; 65% k-NN; 60%

k-NN: k-nearest neighbors, ANN: Artificial neural network, LR: Logistic regression, SVM: Support vector machine

In the studies of Kumar et al. (3), they showed that the handwriting of men and women can be distinguished from each other with the features obtained from the sentences of 200 people. In the same way, Hamid and Loewenthal (8), in their study with writing samples from 30 people, made a sex estimation using 25 experts and showed that this is possible. In addition, in the study of Al Maadeed and Hassaine (4), they stated that they could use the automatic sex handwriting classification system with a certain sentence written to the students. However, Liwicki et al. (15) also used a mathematical model to classify sex with an accuracy of 67.5% in the data they collected. In the study of Riza et al. (19), on the other hand, they estimated the sex with 76% accuracy by using 49 variables with effects such as height, pressure, margin etc. of certain words obtained from 75 people.

These studies show that by looking at people's handwriting, sex can be determined and an automatic system can be established.

In the articles made using the QUWI, MHSH, IAM, KHATT, ICDAR2013 and CEDAR databases, it is aimed to perform sex classification with different structuring in the pattern recognition method (1,3-5,13,15-18,20,23,24).

Success achieved varies between 68.90% and 82%. Using combinations different patterns (LBP, HOG, GLCM, SFTA) ANN, S, DT, KNN and RF analyzes were implemented in best success. Apart from this, any success rate varies between 68.90% and 77%. You must be a good computer scientist for these analyses. It is difficult for these non-expert scientists to perform these analyzes and expert systems are very demanding. In this study, it is aimed to develop an easier, faster and higher classification success method with simpler data.

While determining the belonging in forensic handwriting comparisons, conducting a research on the sex of the writer will both facilitate the comparison and increase the reliability of the conclusion reached. Therefore, developing methods to be used in sex prediction and increasing their accuracy will greatly advantage the forensic document examination society.

In the present study, parts of the same letters in the same words in identical texts written by 50 female and 50 male participants were measured aiming at revealing the differences between them as well as to investigate the success rate in sex determination using data mining.

MATERIALS AND METHODS

In line with the aim of the study, 11 sentences in Turkish containing the letters "b, d, f, g, h, k, p, t, y" at initial, medial, and end positions were written by 50 female and 50 male individuals who were higher education students and/or graduates. Cursive samples were not included in this study. Handwriting samples from people were collected on A4 paper. These A4 papers were scanned as a whole at 300 dpi and saved in jpeg format. Then opened in A4 size in Photoshop. Parts of letters were measured at x300 magnification (Figure1).

The body and extension parts of these letters were measured in millimeters in Adobe Photoshop CS6 by three different researchers, as shown in Figures 1, 2. Whether a statistical difference existed between the extensions of the letters written by male and female participants was evaluated by an Independent sample t-test.

The most commonly used methods for classifying variables in data mining methods are support vector machine (SVM), artificial neural network, logistic regression (LR) analysis, and k-nearest neighbor (k-NN). In different studies, these methods have been reported to be superior to one another according to the type of data used. Twenty-seven different measurements of the letters were obtained at initial, medial, and end positions in the present study. As the number of subjects was 100, variables that showed significant differences by sex were included in the selection of variables. Accordingly, in this study, sex determination was performed based on the stroke lengths of the letters "b, d, h, k, t," which were determined as significant in the independent sample t-test and of the letter "p," which was significant at the medial position, using the aforementioned four methods.

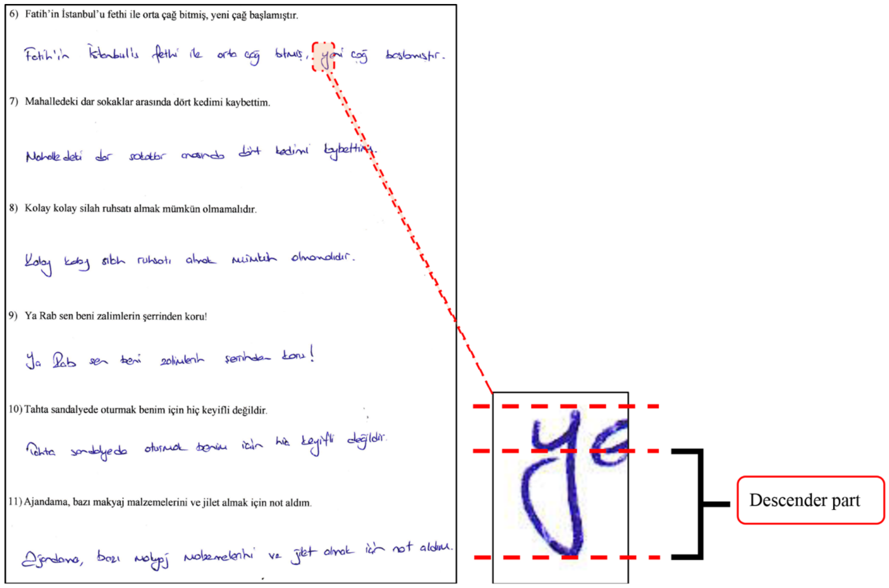

Figure 1. Example from measurement from database

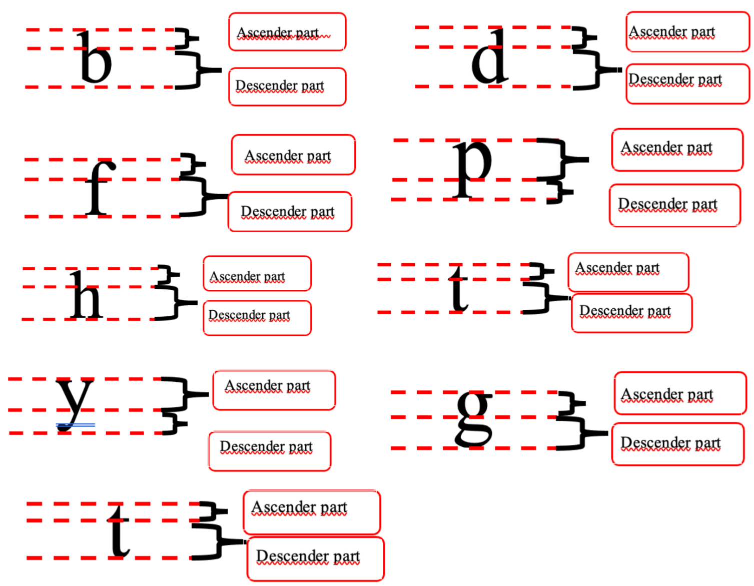

Figure 2. Measured parts of the letters

Support Vector Machine

SVM is a machine learning method developed in the late 1960s by Vladimir Vapnik and Alexey Chervonenkis and is primarily based on statistical learning theory. The SVM method has been used frequently in recent years, especially in data mining, for classification problems in datasets where patterns between variables are unknown (25). This method is basically intended as a linear classifier in solving two-class problems and then generalized to the solution of linearly inseparable or multiclass classification problems, and it has been widely used in the solution of these problems. When applied to linearly separable data, SVM aims to select the line that will make the margin the highest among an infinite number of lines that can separate the data. In the case of linearly inseparable data, SVM transforms the original data into a higher dimensional space with a mapping method and tries to find the linear separating hyperplane that can be optimized to classify the data (26). Models use kernel functions for this purpose. The kernel function of choice affects the performance of the system, and different results can be obtained with different kernel functions.

Below is an illustration of how the SVM method works in a two-dimensional space (Figure 3).

Artificial Neural Networks

ANNs are computer systems inspired by the characteristics of biological nervous systems (information generation, description, estimation, etc.) (29). As in the biological nervous systems, ANNs are formed by a combination of cells. Generally, ANN architecture is defined in three layers: input, intermediate or hidden, and output (27).

More than one intermediate hidden layer can be present in a network. To date, it has not been determined how many hidden layers should be used in an ANN and how many nerve cells should be in each hidden layer. The solution to this situation, which varies according to the problem, has been through trial and error (28,29). A network with several hidden neurons cannot distinguish complex patterns because it can only make linear predictions. Moreover, a large number of hidden neurons prevents the network from generalizing (28,30). Because additional layers exist between the input and output layers in solving non-linear problems, the network architecture becomes multilayered, as shown below (Figure 4).

The backpropagation algorithm is widely used as the learning algorithm of ANN in multilayer feedforward networks.

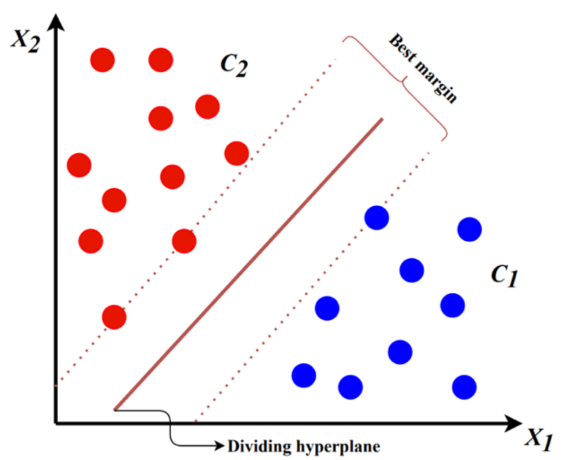

Figure 3. Classification of data using support vector machine

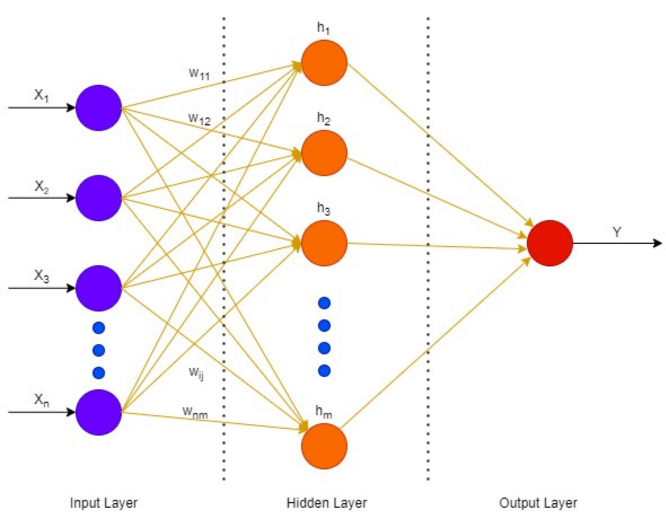

Figure 4. Architecture of a neural network

In backpropagation networks, data is processed from the input layer to the hidden layer and then to the output layer. The purpose in of obtaining values close to the targets as output is to find the optimal weights.

In the present study, Levenberg-Marquardt learning algorithm is used to adjust weights in multilayer feedforward networks. In the hidden and output layers, logistic sigmoid nonlinear function (logsig) and linear transfer function (purelin), respectively, have been used as activation functions.

Logistic Regression

LR analysis is a regression method that helps classification and assignment. In most biological, health, and socio-economic studies conducted to reveal cause and effect relationships, some of the variables examined comprise two-level data such as positive-negative, successful-unsuccessful, or yes-no. In this way, if the dependent variable comprises two-level or multilevel categorical data, LR analysis has an important place in examining the cause-effect relationship between the dependent and independent variables (31). In the LR analysis, one of the purposes of which is classification and the other is to investigate the relationships between dependent and independent variables, the dependent variable takes categorical values. Independent variables can be continuous or categorical variables. Besides being applicable when the dependent variable is a two-level variable, such as 0 or 1, or a discrete variable with more than two levels, its mathematical flexibility and easy interpretability increase the interest in this method (32).

K-nearest Neighbor (k-NN) Algorithm

K-NN algorithm, T. M. Cover ve P. E. K-NN algorithm, proposed by Hart, is a classification method by which the nearest neighbor of the class in which the sample data point is present is determined according to the k value (33). This algorithm is one of the best-known, old, simple, and effective pattern classification methods and is popularly used among machine learning algorithms (34). Classification of objects is an important research area and is applied in a wide variety of fields such as pattern recognition, data mining, artificial intelligence, statistics, cognitive psychology, medicine, and bioinformatics (35). The k-NN algorithm is especially preferred in classification applications owing to its advantages such as easy applicability and resistance to noisy training data. Despite these advantages, it also brings some disadvantages such as processing load increases with the number of datasets and variables' performance being affected by parameters or features such as the number of neighbors, distance criteria, and the number of variables (33). k-NN calculates the probability of data that is considered to belong to the class of its neighbors according to the status of its closest neighbor (25). The following figure shows to which class the data indicated with an asterisk will belong in cases k=3 and 6 (Figure 5).

Classification Criteria

The confusion matrix evaluates the performance of classification models and tells us how good our classification model is while making predictions on test data. The rows of the matrix contain actual values whereas the columns contain predicted values. Predicted values are values calculated by the model, and actual values are true values for the given observations. With the help of the confusion matrix, different parameters such as accuracy and precision can be calculated for the model. These values indicate how effective the used methods are. True positive (TP), true negative (TN), false positive (FP), and false-negative (FN) values in the confusion matrix are used to calculate the following values. Because our aim in this study was to correctly predict females, the correct prediction of females in the confusion matrix was TP, and the correct prediction of males was FP; moreover, the incorrect prediction of females was TN, and the incorrect prediction of males was FN.

In this study, accuracy (ACC), error rate (ERR), precision (PREC), sensitivity (SENS), specificity (SPEC), F-measure (FM), Youden's index (YI), kappa (κ) statistics, true negative rate (TPR), false positive rate (FPR), and receiver operating characteristic (ROC) area values were used.

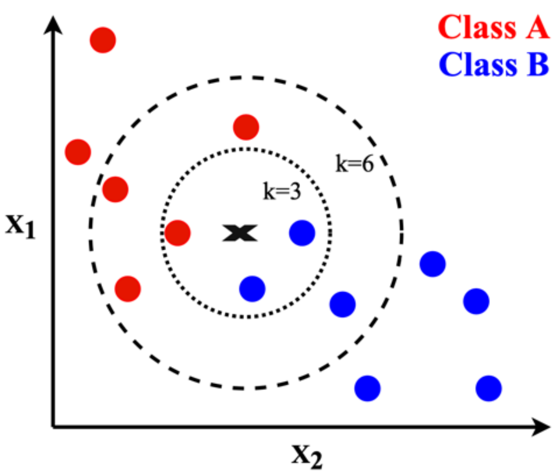

Figure 5. The k-nearest neighbor prediction with k=3 and 6

RESULTS

Statistical Analyses

The data collected from 100 different people were measured by three different people, and measurement error was examined using the Friedman analysis. Consequently, no statistically significant difference was observed among the measurements performed by the three different researchers. Therefore, the analysis was carried on with the data measured by one person. Extensions of the letters "b, d, h, k, t" were found to be statistically significantly longer in males. No such difference was found in the letters "p," "f," "y," and "g" (Table 2).

Table 2. Statistical values of the examined letters belonging to male and female participants

Letter	Gender	Initial		Medial		End	
		x̄	std	x̄	std	x̄	std
k	Female	1.32*	0.62	1.39*	0.65	1.33*	0.64
	Male	1.97*	0.82	1.99*	0.95	2.08*	0.85
p	Female	1.96	0.51	1.94*	0.61	2.16	0.61
	Male	2.19	0.67	2.26*	0.85	2.42	0.87
f	Female	3.25	0.92	2.86	1.00	3.13	1.05
	Male	3.58	1.13	2.99	0.91	3.47	1.14
t	Female	2.12*	0.66	1.82*	0.75	1.90*	0.63
	Male	2.59*	1.05	2.41*	1.03	2.41*	1.00
b	Female	1.70*	0.54	1.69*	0.72	1.85*	0.67
	Male	2.16*	1.03	2.27*	1.03	2.43*	1.11
d	Female	1.97*	0.73	1.79*	0.76	1.77*	0.82
	Male	2.73*	1.06	2.47*	1.08	2.72*	0.95
h	Female	1.58*	0.59	1.63*	0.67	1.73*	0.68
	Male	2.16*	0.88	2.26*	1.09	2.24*	0.95
g	Female	3.27	1.06	2.88	0.97	3.05	1.16
	Male	3.37	1.34	2.79	1.00	2.95	1.06
y	Female	2.96	1.29	3.15	1.23	3.14	1.11
	Male	3.28	1.14	3.09	1.19	3.45	1.36

*p<0.001, std: Standard

Here, 9 letters have a total of 27 measurements, including the initially middle and end. This means 27 different variables. In prediction models; If the number of variables is high and the number of samples is few, "overfitting" occurs. To avoid this, one of the variable selection methods such as random forest, chaid analysis or variable reduction methods such as principal component analysis can be used. Here, the independent sample t-test was preferred because of the grouping. When the sample size of the groups was 50 and the alpha value was 0.05, the effect size was found to be 0.80 when examined. This effect size is sufficient. For this, the G power software was used. Accordingly, it is seen that the number of samples used in the study is sufficient. A prediction model has been established with the data we have. Thanks to this model, it can be determined with 82% accuracy whether a person writing the same words is male or female.

Data Mining Analyses

With the inclusion of letter "p," which was determined as significant at the medial position, to the letters "b, d, f, h, k, t"; ANN, k-NN, SVM, and LR were applied, and the most successful result was obtained with ANN (Tables 3 and 4). The analyses were conducted with five cross-validations. The confusion matrix obtained as a result of the analyses is shown in Table 2.

With ANN, 43 out of 50 females were correctly predicted as females, whereas 39 of the males were correctly predicted as males (Table 3). The positive predictive value for ANN was 0.86, and the negative predictive value was 0.78. The analysis with the least error was ANN with an error ratio of 18%. ACC value for ANN was 0.82. The highest Kappa value was obtained with 0.64 in the ANN analysis. The LR (+) value obtained as a result of the analysis was best in the ANN analysis with 5.23. Considering the ROC test, the best explanation was approached with the ANN analysis. Consequently, there is an 88% probability of determining a person as a male or a female when the measurement values of that person are known (Table 4).

Correct Classification Rate (Accuracy): The closer the correct classification rate to 1, the higher the performance of the test. When this value is below 0.50, the classification made by the test performed can be said to be by chance. Herein, the highest correct classification rate was in ANN (82%), and the lowest was in the KNN analysis (60%).

Kappa Coefficient: It is a coefficient that provides information about reliability by correcting chance agreement that occurs solely by chance. This coefficient takes a value between 0 and 1. A value of 0-0.39 implies poor agreement, 0.40-0.75 good agreement, and 0.76-1.00 perfect agreement. Herein, the Kappa value varied between 0.20 and 0.64. The analysis with the lowest agreement was KNN, and the analysis with the highest was ANN. Accordingly, a good level of agreement existed in the assignments performed using the ANN and SVM analyses. Conversely, a weak agreement existed in the assignments performed using KNN and LR.

Table 3. Data mining analysis results of female and male participants

	KNN		Logistics		SVM		ANN	
	Female	Male	Female	Male	Female	Male	Female	Male
Female	31	19	32	18	37	13	43	7
Male	21	29	17	33	16	34	11	39

KNN: k-nearest neighbor, SVM: Support vector machine, ANN: Artificial neural network

Table 4. Comparison criteria at the end of the analyses

	Logistics	KNN	SVM	ANN
ACC	0.65	0.60	0.71	0.82
ROC	0.69	0.60	0.77	0.88
Kappa	0.30	0.20	0.42	0.64
F1	0.65	0.61	0.72	0.83
Sensitivity	0.65	0.60	0.70	0.80
Specificity	0.65	0.60	0.72	0.85
Precision	0.64	0.62	0.74	0.86
Negative prediction rate	0.66	0.58	0.68	0.78
ERR	0.35	0.40	0.29	0.18
Youden index	0.30	0.20	0.42	0.64
OR (+)	1.85	1.51	2.52	5.23
OR (-)	0.54	0.67	0.42	0.24

ACC: Accuracy, ROC: Receiver operating characteristic, KNN: k-nearest neighbor, SVM: Support vector machine, ANN: Artificial neural network, ERR: Error rate, OR: Odds ratio

Likelihood Ratio: Two likelihood ratios, positive and negative, exist. LR (+) indicates the number of correct positives the model gives versus an FP. The higher this ratio, the better it distinguishes the positive state. LR (−) defines the number of false negatives a test gives for each TN. The smaller this ratio (closer to zero), the better the negativity success of the test. The fact that it is equal to 1 in both the likelihood ratios indicates the situation where the test is the most unsuccessful. The LR (+) value obtained as a result of the analysis was best in the ANN analysis with 5.23. SVM, LR, and KNN followed ANN, respectively. The worst result was obtained in the KNN analysis. The smallest LR (−) value was obtained in the ANN analysis. Similarly, SVM, LR, and KNN followed ANN, respectively.

Youden Index: It gives an overall assessment of the performance of the test and is used for comparing multiple tests. The test result is desired to be close to 1.

In the study, Kappa coefficient and Youden index values LR, KNN, SVM and ANN were found to be 0.30, 0.20, 0.42, 0.64, respectively. Although the values are the same, these analyzes are included in the study because there are differences in terms of interpretation.

It is sometimes difficult in a study to decide which of the methods used had the best performance because while the sensitivity analysis result is high in some methods, the specificity rate may be high in others. For this reason, combined criteria such as correct classification rate, LR, and odds ratio, which are obtained by combining sensitivity and specificity values, are used to compare the performance of different methods. The results are as follows.

Sensitivity shows the percentage of actual positives identified as positive in the results of the developed test. The ratio obtained as a result of the test is desired to be close to 1. Among the performed analyses, the ANN analysis identified actual females as females with 80% success. The success of the KNN analysis was 60%.

Specificity indicates the percentage of actual negatives identified as negative in the results of the developed test. The ratio obtained as a result of the test is desired to be close to 1. As with sensitivity, the ANN analysis achieved the most successful results in specificity as well predicting actual males as males with 85% success.

Positive predictive value gives the correct prediction of women, and it was 0.62 with KNN, 0.64 with LR, 0.74 with SVM, and 0.86 with ANN.

The negative predictive value gives the probability of negative assignments made by the applied test being actual negatives and is desired to be close to 1. As a result of the analysis, correct classification ratios were 0.58 with KNN, 0.66 with Logistics, 0.68 with SVM, and 0.78 with ANN.

Error, which is the rate of incorrect classification, was the highest in KNN at 40%, and the lowest in the ANN method at 18%.

The optimal cut-off point values that distinguish male and female status can be determined by the ROC curve analysis (36). With ROC analysis, the correct prediction rate of a test is measured by the area under the curve (AUC). The AUC value indicates the overall accuracy of the test. The values of 0.90-1.00 indicate perfect accuracy, 0.80-0.90 good accuracy, 0.70-0.80 moderate accuracy, 0.60-0.70 poor accuracy, and below 0.60 imply that the test is not useful (37).

The ROC curve (graph) is obtained connecting the sensitivity results obtained according to all cut-off values marked on the y-axis and the specificity results marked on the x-axis. The value, denoted as AUC at the end of the analysis, represents the "area under the curve," and the determination value increases as it approaches 1. The AUC value obtained as a result of the analysis varied between 0.60 and 0.88. Herein, the best explanation was approached with the ANN analysis. Consequently, when the measurement values of a person are known, there is an 88% probability of determining the correct sex of that person.

DISCUSSION

In forensic handwriting examinations, if more than one person is examined, the ability to determine the sex of the writer is of great benefit. This situation will contribute to the reliability of the study by acting as an additional determination factor and thus will reduce the workload by eliminating individuals of different sexs and will set the ground for faster results. For this purpose, classifying the handwritings of males and females is necessary. Several studies exist on this subject in the literature (1,5,13,18,23,36,38-40). In these studies, the probabilities of handwriting belonging to a female or male participant were determined using various texts and measurement techniques. Generally, this rate remains at approximately 70%.

In the present study, 50 female and 50 male individuals were asked to write 11 sentences in the Turkish language containing the letters "b, d, f, g, h, k, p, t, y." In the examination, it was determined that the ascender and descender parts of the letters "b, d, h, and k" were statistically significantly higher in males than in females. No such difference was detected in the letters "p," "f," "y," and "g." The extensions of these two types of letters are made considerably long in both males and females compared to their bodies.

The measurement process was conducted by three different researchers, and no statistically significant difference was found among their measurements, which shows that the measurement process is repeatable.

The data obtained in the measurement of ascender and descender parts were used to determine the probability of handwriting belonging to a female or male participant using the data mining techniques, i.e., KNN, SVM, LR, and ANN. Moreover, the method that made the best predictions was also investigated.

Although the success rate in studies on sex prediction using SVM varies between 48.9% and 77.98% (5,13,16,23,38,40), the correct classification rate obtained in this study was 71%.

Despite the success rate in studies on sex prediction with ANN varies between 55% and 74.7% (1,5,13,39), a correct classification rate of 82% was obtained in the present study.

In the sex prediction study using global features conducted by Ibrahim et al. (22), the ROC value obtained from the feature type with the highest accuracy value was found to be 0.658 whereas the ROC value in the prediction made using local properties was 0.534. In the analyses performed in the present study, the ROC was found to be 0.88 with ANN and 0.77 with SVM.

The success rate was reported to be 74% in a study conducted by Sesa-Nogueras et al. (21) to predict sex with dynamic features in handwritings written on tablets. Similarly, in the study conducted by Liwicki et al. (15) on handwritings collected online using different SVM types, the maximum correct classification rate obtained was 62.9%, which is approximately 10% below that obtained in the present study. It has been reported by Erbilek et al. (38) that sex classification was performed with 75% accuracy through handwritings collected online using the SVM classifier.

Study Limitations

One of the limitations of this study is that it has printed text. Additional work is needed for cursive text. In this way, it can be evaluated whether there is a difference between cursive and printed letters.

One of the study's limitations is that the slant and other spatial-geometric features were not included in the analysis.

For this reason, research on sex estimation is carried out in Handwriting examinations.

It will be helpful to compare the results by repeating similar studies, taking into account the slope and without taking into account the slope in future studies.

Another limitation is the precision of the measurement technique. In this study, a model was established with LR, SVM, k-NN and ANN methods. Among these methods, ANN came to the fore with the best prediction rate. The validity of the ANN model can be determined by comparing it with the words of known sexs to be obtained from a Forensic case. Since this study was in an experimental environment, the model was run on the same words. In this study, it was not investigated whether the correct classification was made by having the same people write different words with the letters used in the model. This can be examined in another study.

The success rate of the method used in the present study will further increase with the inclusion of additional dynamic features in handwritings written on tablets.

CONCLUSION

In this study accuracy rates are 0.65, 0.60, 0.71 and 0.82 for Logistics, KNN, SVM and ANN respectively. The results showed that the model developed using ANNs achieved significant success in sex prediction. The results obtained from this study were higher than those obtained in other studies in the literature. The biggest difference of this study from other studies is that it can predict with higher accuracy without pattern recognition; however, not all prior studies employed pattern recognition. The high accuracy rates achieved in this study without pattern recognition indicate that better rates will be achieved when pattern recognition is used. In other studies with artificial neural networks, databases such as IAM and KHATT are generally used and the accuracy rate varies between 60-80%. In this study, 82% accurate classification rate was obtained faster with a simpler measurement method. In our study, 9 letters, 27 variables (initially, in the middle, at the end), which are the extensions, were selected by choosing those that differ according to sex. For this reason, it showed a better performance than the others. High accuracy rates without pattern recognition in this study indicate that higher rates will be achieved when pattern recognition is used. In future studies, a more detailed distinction can be made by adding variables such as age and hand used to the model. Based on studies using similar methods, the result we obtained is promising but in need of improvement for its application to forensic cases.

ETHICS

Ethics Committee Approval: Approval for the current study was granted by the Istanbul University-Cerrahpaşa, Social and Human Sciences Ethics Committee (approval no: 2019/71, dated: 15.11.2019).

Authorship Contributions

Concept: G.Ç., Design: D.Ö.K., G.Ç., Data Collection or Processing: D.Ö.K., Y.K., T.Ü.K., Ö.K., İ.D., Analysis or Interpretation: D.Ö.K., Y.K., Ö.K., İ.D., G.Ç., Literature Search: D.Ö.K., Y.K., T.Ü.K., İ.D., Writing: D.Ö.K., Ö.K., İ.D., G.Ç.

Conflict of Interest: No conflict of interest was declared by the authors.

Financial Disclosure: The authors declared that this study received no financial support.

REFERENCES

1. Bandi KR, Srihari SN. Writer demographic classification using bagging and boosting. In Proceedings of the twelfth International Graphonomics Society Conference. 2005; 26-29; Salerno, Italy, 2005;133-137.
2. Huber RA, Headrick AM. Editors. Handwriting Identification: Facts and Fundamentals. Science, Scientific Method, and Writing Identifications. Florida, Boca Roton: CRC Press LLC. 1999;362-398.

3. Kumar S, Saran V, Vaid, BA, Gupta AK. Handwriting and gender: A statistical study. Z Zagadnień Nauk Sądowych, 2013;95:620-626.
4. Al Maadeed S, Hassaine A. Automatic prediction of age, gender, and nationality in offline handwriting. J Image Video Proc. 2014;10:1-10. https://doi.org/10.1186/1687-5281-2014-10
5. Akbari Y, Nouri K, Sadri J, Djeddi C, Siddiqi I. Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata. Image and Vision Computing. 2017;59:17-30. https://doi.org/10.1016/j.imavis.2016.11.017
6. Goodenough FL. Sex differences in judging the sex of handwriting. J Soc Psychol. 1945;22(1):61-68. http://doi.org/10.1080/00224545.1945.9714182
7. Marzinotto G, Nunez JCR, Yacoubi ME, Garcia-Salicetti S. Age and Gender Characterization through a Two Layer Clustering of Online Handwriting. In: Proceeding of sixteenth International Conference on Advanced Concepts for Intelligent Vision Systems. ACIVS; 2015 26-29 Oct; Catania, Italy, 2015;428-439. http://doi.org/10.1007/978-3-319-25903-1_37.
8. Hamid S, Loewenthal KM. Inferring gender from handwriting in Urdu and English. J Soc Psychol. 1996;136(6):778-782. http://doi.org/10.1080/00224545.1996.9712254
9. Binet A. Lea Efivelations de 1 'Ecriture d 'Apres un Controle Scientiftque, Paris. 1906:1-22.
10. Young PT. Sex differences in handwriting. J Appl Psychol. 1931;15(5):486-498. http://doi.org/10.1037/h0072627
11. Tomai CI, Kshirsagar DM, Srihari SN. Group discriminatory power of handwritten characters. In: proceeding of the seventeeth ICPR; 2004 26 Aug; Cambridge, UK, 2004;638-641. http://doi.org/10.1109/ICPR.2004.1334329
12. Topaloglu M, Ekmekci S. Gender detection and identifying one's handwriting with handwriting analysis. Expert Syst Appl. 2017;79(1):236-243. http://doi.org/10.1016/j.eswa.2017.03.001
13. Siddiqi I, Djeddi C. Raza A, Souici-Meslati L. Automatic analysis of handwriting for gender classification. Pattern Anal Appl 2015;18(4):887-899. http://doi.org/10.1007/s10044-014-0371-0
14. Sokic E, Salihbegovic A. Ahic-Djokic M. Analysis of off-line handwritten text samples of different gender using shape descriptors. In: proceeding of the nineteenth International Symposium on Telecommunications (BIHTEL); 2012 Oct 25-27; Sarajevo, Bosnia and Herzegovina, IEEE publications 2012; 1-6.
15. Liwicki M, Schlapbach A, Loretan P, Bunke H. Automatic detection of gender and handedness from on-line handwriting. In: Proceeding of the thirteenth Conference of the International Graphonomics Society; 2007 Nov 11-14; Melbourne, Australia, 2007;179-183.
16. Ahmed M, Rasool AG, Afzal H, Siddiqi I. Improving handwriting based gender classification using ensemble classifiers. Expert Syst Appl. 2017;85(1):158-168. http://doi.org/10.1016/j.eswa.2017.05.033
17. Morera Á, Sánchez Á, Vélez JF, Moreno AB. Gender and handedness prediction from offline handwriting using convolutional neural networks. Complexity, 2018. http://doi.org/10.1155/2018/3891624
18. Youssef AE, Ibrahim AS, Abbott AL. Automated gender identification for Arabic and English handwriting. In: Proceeding of fifth International Conference on Imaging for Crime Detection and Prevention (ICDP); 2013 16-17 Dec; London, UK, IET publications 2013;1-6. http://doi.org/10.1049/ic.2013.0274
19. Riza LS, Zainafif A, Rasim SN. Fuzzy rule-based classification systems for the gender prediction from handwriting. Telkomnika, 2018;16(6):2725-2732. http://doi.org/10.12928/telkomnika.v16i6.9478
20. Cha SH, Srihari SN. A priori algorithm for sub-category classification analysis of handwriting. In: Proceedings of Sixth International Conference on Document Analysis and Recognition; 2001 Sep 13-13; Seattle, WA, USA, IEEE publications 2002;1022-1025. http://doi.org/10.1109/ICDAR.2001.953940
21. Sesa-Nogueras, E, Faundez-Zanuy, M, Roure-Alcobé, J. Gender classification by means of online uppercase handwriting: a text-dependent allographic approach. Cogn Comput. 2016;8(1):15-29. http://doi.org/10.1007/s12559-015-9332-1
22. Ibrahim AS, Youssef AE, Abbott AL. Global vs. local features for gender identification using Arabic and English handwriting. In: Proceeding of International Symposium on Signal Processing and Information Technology (ISSPIT). 2014:15-17 Dec; Noida, India, IEEE publications 2015;155-160. http://doi.org/10.1109/ISSPIT.2014.7300580.
23. Bouadjenek N, Nemmour H, Chibani Y. Age, gender and handedness prediction from handwriting using gradient features. In: Proceeding of thirteenth International Conference on Document Analysis and Recognition (ICDAR);2015 Aug 23-26; Tunis, Tunisia, IEEE publications 2015;116-1120. http://doi.org/10.1109/ICDAR.2015.7333934
24. Guerbai Y, Chibani Y, Hadjadji B. Handwriting gender recognition system based on the one-class support vector machines. In: Proceeding of Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA); 2017 28 Nov- 1 Dec; Montreal, QC, Canada, IEEE publications 2018; 1-5. http://doi.org/10.1109/IPTA.2017.8310136
25. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. J Biomed Inform. 2002;35(5-6):352-359. https://doi: 10.1016/S1532-0464(03)00034-0
26. Noble, WS. What is a Support vector machine? Nat Biotechnol. 2006;24(12):1564-1567. http://doi.org/10.1038/nbt1206-1565
27. Skapura DM. "Building Neural Networks" Addison-Wesley, New York; 1996; 29-64 (ch 2).
28. Haykin S. Neural Networks: A Comprehensive Foundation" MacMillan. New York. 2008;22-24 (ch:1).
29. Oztemel E. Artificial neural networks. Papatya Publishing. Istanbul. 2003;29-31 (ch:2) (Turkish Translate)
30. Chaudhuri BB, Bhattacharya U. Efficient Training and Improved Performance of Multilayer Perceptron in Pattern Classification. Neurocomputing. 2000;34(1-4):11-27. http://doi.org/10.1016/S0925-2312(00)00305-2
31. Agresti A. An Introduction to Categorical Data Analysis. Logistic regression. John Wiley and Sons. Inc., 2019;89-92 (ch:4), Third Edition.
32. Lemeshow S, Hosmer D. Applied Logistic Regression (Wiley Series in Probability and Statistics). The Multiple Logistic Regression Model. Wiley-Interscience; 2013; 35-36 (ch:2) Third Edition.
33. Bhatia N, Vandana. Survey of nearest neighbor techniques, IJCSIS. 2010;8(2):302-305.
34. Qiu XY, Kang K, Zhang HX. Selection of kernel parameters for K-NN. In: Proceeding of International Joint Conference on Neural Networks (IJCNN), 2008 1-8 June; Hong Kong, China, IEEE publications. 2008;61-65. http://doi.org/10.1109/IJCNN.2008.4633767
35. Batista Gustavo. EAPA, Silva, DF. How k-nearest neighbor parameters affect its performance. In: Proceeding tenth Argentine Symposium on Artificial Intelligence (ASAI), 2009 24-25 Aug; Mar Del Plata, Argentina. 2009;95-106.
36. Hassaïne A, Al Maadeed S, Aljaam J, Jaoua A. competition on gender prediction from handwriting. In: Proceeding of twelfth International Conference on Document Analysis and Recognition (ICDAR). 2013:25-28 Aug; Washington, DC, USA, IEEE publications 2013;1417-1421. http://doi.org/10.1109/ICDAR.2013.286
37. Demir I. Statistics Guide with SPSS. Regression analysis for categorical data. Istanbul 2020;438-439. (ch:14) (Turkish Translate).
38. Erbilek M, Fairhurst M, Li C. Exploring gender prediction from digital handwriting. In: Proceeding of twenty forth Signal Processing and Communication Application Conference (SIU). 2016;16-19 May; Zonguldak, Turkey, IEEE publications 2016;789-792. http://doi.org/10.1109/SIU.2016.7495858
39. Mirza A, Moetesum M, Siddiqi I, Djeddi C. Gender classification from offline handwriting images using textural features. In: Proceeding fifteenth International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016 23-26 Oct; Shenzhen, China, IEEE publications. 2017;395-398. http://doi.org/10.1109/ICFHR.2016.0080
40. Gattal A, Djeddi C ,Siddiqi I, Chibani Y. Gender classification from offline multi-script handwriting images using oriented basic image features (oBIFs). Expert Syst Appl. 2018;99(1):155-167. http://doi.org/10.1016/j.eswa.2018.01.038