SAUCIS

Sakarya University Journal of Computer and Information Sciences

http://saucis.sakarya.edu.tr/

SAKARYA UNIVERSITY

DOI:10.35377/saucis...1337649

RESEARCH ARTICLE

A Novel Gender Classification Model based on Convolutional Neural Network through Handwritten Text and Numeral

Pakize Erdoğmuş[1,], Abdullah Talha Kabakuş[1], Enver Küçükkülahlı[1], Büşra Takgil[1], Ezgi Kara Timuçin[1]

[1] Duzce University, Faculty of Engineering, Department of Computer Engineering, Düzce, Türkiye

Corresponding author:
Pakize Erdoğmuş, Düzce University, Faculty of Engineering, Department of Computer Engineering, Düzce, Türkiye
E-mail address:
pakizeerdogmus@duzce.edu.tr

Submitted: 04 August 2023
Revision Requested: 06 September 2023
Last Revision Received: 10 September 2023
Accepted: 28 September 2023
Published Online: 30 September 2023

Citation: Erdogmus P. et al. (2023). A Novel Gender Classification Model based on Convolutional Neural Network through Handwritten Text and Numeral. *Sakarya University Journal of Computer and Information Sciences*. 6 (3)
https://doi.org/10.35377/saucis...1337649

ABSTRACT

Human handwriting is used to investigate human characteristics in various applications, including but not limited to biometric authentication, personality profiling, historical document analysis, and forensic investigations. Gender is one of the most distinguishing characteristics of human beings. From this point forth, we propose a novel end-to-end model based on Convolutional Neural Network (CNN) that automatically extracts features from a given handwritten sample, which contains both handwritten text and numerals unlike the related work that uses only handwritten text and classifies its owner's gender. In addition to proposing a novel model, we introduce a new dataset that consists of 530 gender-labeled Turkish handwritten samples since, to the best of our knowledge, there does not exist a public gender-labeled Turkish handwriting dataset. Following an exhaustive process of hyperparameter optimization, the proposed CNN featured the most optimal hyperparameters and was both trained and evaluated on this dataset. According to the experimental result, the proposed novel model obtained an accuracy as high as 74.46%, which overperformed the state-of-the-art baselines and is promising on such a task that even humans could not have achieved highly-accurate results for, as of yet.

Keywords: Handwriting, gender classification, convolutional neural network, computer vision, forensic science

1. Introduction

Today's rapidly advancing technological world does indeed bring forth numerous security challenges. These technological advancements serve not only benign consumers but also, unfortunately, enable malicious entities. Authentication technologies are employed to address security concerns by verifying or recognizing a person's identity through various factors, including passwords and facial features. Biometric authentication is one of the authentication methods using inherence factors such as fingerprints, DNA (DeoxyriboNucleic Acid), face, or retina. The utilization of computers for human recognition based on physical and behavioral traits traces its origins to the digital computer revolution of the 1960s [1]. But even today, after more than 60 years, biometric studies remain fresh since new technologies require using more secure applications. Gender detection, the ability to detect an individual's gender based on distinct physiological features and patterns, is also one of the biometric factors for authentication and forensic applications. In addition to this, human handwriting is used to investigate human characteristics in various applications, including but not limited to personality profiling [2], [3], historical document analysis [4], and forensic investigations [5], [6]. In areas like artificial intelligence (AI), computer vision, and human-computer interaction, precise gender detection becomes pivotal in crafting personalized and inclusive user experiences. Furthermore, gender detection can play a vital role in promoting fair representation and addressing biases, particularly within domains such as criminal justice and employment. For criminal justice, gender detection provides, including but not limited to (i) equitable treatment, (ii) risk assessment, (iii) victim identification, (iv) investigation and profiling, and (v) evidence handling. Gender stands as the primary physiological and physical distinction among individuals. It influences how people perceive themselves and each other [7]. So, from hand use to brain functions, there are fundamental differences between genders [8]– [10]. The reasons behind its importance for human distinction can be listed as follows: (i) social identity and cultural context, (ii) self-perception and identity formation, (iii) social interaction and communication, (iv) access to opportunities, (v) representation, (vi) advocacy, (vii) psychological well-being, and $(viii)$ historical significance.

Drawing is the oldest communication tool, including the alphabet, digits, and traffic signs and has been used for centuries by various cultures as a means to convey ideas, stories, and information visually. A drawing consists of a lot of information about its creator, such as graphology, expressiveness, attention to detail, creativity, imagination, emotions, feelings, cultural and social influences, and communication style. Even having *Alzheimer's* or *Parkinson's disease* [11] is one of these conveyed information. Drawing differs between genders. Sex differences in drawings have been discussed by a number of researchers [12], [13]. Handwriting, as a biometric modality, offers an unobtrusive means of inferring gender-related attributes without the need for direct personal interaction. This capacity has sparked interest in a wide range of fields, including but not limited to psychology, forensics, linguistics, and AI. Examining the connection between handwriting patterns and the identification of gender offers valuable insights into how individuals encapsulate their identities within the very act of writing. The motivation behind this study is to explore the potential presence of gender differences in handwriting drawings. When we have reviewed the literature, we have found that there exist some studies on this topic [14]–[17].

In this paper, we explore the relationship between handwriting and gender. In other words, we classify an individual's gender through handwriting. In order to verify the relation between handwriting and gender, we collected both handwritten Turkish sentences consisting of the whole letters in the Turkish alphabet and the numbers from zero to nine. After collecting the raw data, we digitized it and constructed a dataset for both handwritten numbers and text. Some image pre-processing methods, such as denoising and morphological operations, have been applied to the dataset. After the pre-processing of the dataset, we have proposed a novel Convolutional Neural Network (CNN) model that accepts the handwritten text and numerals as the input and generates the detected gender, which covers two classes, namely, (i) *male*, and (ii) *female*. The deliberate design choice for the proposed model was to adopt a CNN structure, given that CNNs have consistently demonstrated state-of-the-art performance across a spectrum of computer vision tasks. These tasks encompass a wide range, such as image classification, object detection, object tracking, medical image analysis, autonomous driving, facial recognition, and document analysis, among others. The main contributions of this article can be summarized as follows:

- We propose a novel CNN model combining both handwritten text and numerals features as input. To the best of our knowledge, this is the first study that makes gender classification through a combination of handwritten text and numerals.
- Thanks to proposing a model based on DNN, neither manual feature extraction nor manual feature selection was required. Instead, an end-to-end solution was proposed.
- Since there does not exist a public gender-labeled Turkish handwritten dataset, we introduce a new dataset that consists of 530 gender-labeled Turkish handwritten samples as another contribution to the research field.
- A comprehensive range of values for each hyperparameter was evaluated in automated manner to discover the most optimal combination of hyperparameters that yield the highest classification performance for the proposed model.
- A wide range of widely used traditional Machine Learning (ML), and DNN models was employed as the baseline of the proposed CNN model. According to the experimental result, the proposed CNN model that yields both handwritten text and numerals obtained better accuracy, an accuracy as high as 74.46%, than the state-of-the-art baselines on a task that even humans could not have achieved highly-accurate results for, as of yet [18].

The remaining of the paper is organized as follows: Section 2 briefly reviews the related work. Section 3 outlines the materials and methods employed in this study. In Section 4, we present the experimental results and engage in discussions surrounding them. Lastly, Section 5 encapsulates the paper by drawing conclusions and suggesting potential avenues for future exploration.

2. Related Work

There exist studies that deal with gender classification through English, French, and Arabic text, while studies dealing with gender classification for handwritten Turkish text lack in the research field. To the best of our knowledge, there does not exist a study that makes gender classification from handwritten Turkish text. In an early study, *Koppel et al.* [19] introduced a method rooted in a variant of Exponential Gradient for gender classification using documents sourced from the British National Corpus (BNC). Each individual document extracted was delineated by a feature vector encapsulating distinctive characteristics. The dimensionality of these feature vectors was reduced by the elimination of the irrelevant features. According to the experimental result, the proposed model obtained an accuracy of approximately 80%.

Liwicki et al. [20] proposed a model for detecting gender and handedness from online handwriting. In terms of gender detection, they covered two classes, namely, (i) *male*, and (ii) *female*. Regarding handedness detection, they covered left- and right-handedness. To this end, they employed two models: (i) The proposed first model employed *Support Vector Machine* (*SVM*), and (ii) the other model employed *Gaussian Mixture Model* (*GMM*). These proposed models were trained and evaluated on the *IAM-OnDB*, an English handwriting dataset consisting of more than 200 writers with eight handwritten texts per writer which were acquired from a whiteboard. Despite that, the authors used only 100 of them for the training of the gender classifier and 30 of them for the training of the handedness classifier. The gender detection model was evaluated on a set of 50 writers. The obtained accuracy values on this subset were 67.06% and 62.19% when *GMM* and *SVM* were employed for the classification, respectively. The handedness classification model was evaluated on a set of 30 writers. The obtained accuracy values on this subset were 62.57% and 84.66% when *GMM* and *SVM* were employed for the classification, respectively. The limitations of this study are as follows: (i) Both classifiers were trained on a small dataset despite having a relatively large dataset, and (ii) more complex ML models such as Deep Neural Networks (DNNs) were not employed in addition to the employed traditional ML models.

Gattal et al. [21] proposed a handwriting analysis-based gender classification model using *Cloud of Line Distribution* (*COLD*) and Hinge features, which were coupled with two SVM classifiers. The proposed model was evaluated on a subset of the *QUWI* dataset, which consisted of 1,000 samples. The constructed subset was split as follows: 500 samples were used for the training, 250 samples were used for the validation, and the remaining 250 samples were used for the testing. The proposed model obtained an accuracy of 73.60% on the test set.

Morera et al. [22] introduced a CNN-based model for gender and handedness classification. This model was applied to two publicly available handwriting datasets: (i) the *IAM* dataset, containing English text, and (ii) the *KHATT* dataset, containing Arabic text. The experimental findings revealed that the proposed model achieved an accuracy of 80.72% for gender classification on the *IAM* dataset. As for the *KHATT* dataset, the accuracy of the proposed model was calculated at 68.90%.

Rabaev et al. [23] proposed a DNN for gender classification from handwriting images. In this study, they investigated cross-domain transfer learning with *ImageNet* [24] pre-training. The experiments were carried out on two datasets, namely, (i) the *QUWI* dataset, and (ii) a new dataset of documents in Hebrew script. They experimented with various DNNs and demonstrated that advanced DNNs outperformed traditional ML algorithms.

Al Maadeed and *Hassaine* [25] proposed a gender classification approach from offline documents using two approaches as follows: (i) All subjects wrote the same text, and (ii) each subject wrote a different text. They extracted a number of features based on shape, including but not limited to curvatures, chain codes, and stroke orientations. They employed two classification algorithms, namely, (i) *Random Forest*, and (ii) *Kernel Discriminant Analysis* (*KDA*). The proposed model underwent evaluation on the *QUWI* dataset through a series of diverse experiments involving Arabic texts, English texts, and a merged amalgamation of the two. According to the experimental result, an accuracy of 72.30%, was obtained when the proposed model employed *KDA* when documents from both languages were combined and the handwritten texts of subjects were the same.

Bouadjenek et al. [26] introduced a gender classification methodology utilizing handwriting samples. Their approach was founded on a fusion of *Histogram of Oriented Gradients (HOG)* and *SVM*. While *HOG* was employed to extract relevant features, *SVM* was leveraged for classification purposes. This proposed model underwent evaluation using two distinct datasets: (i) the *IAM* dataset and (ii) the *KHATT* dataset. Based on experimental findings, the proposed model achieved a precision of 75.45% on the *IAM* dataset and a precision of 68.89% on the *KHATT* dataset.

Siddiqi et al. [27] proposed a gender classification of handwriting based on slant/orientation, roundedness/curvature, neatness/legibility, and writing texture features. They employed ANN and *SVM* for the classification. The proposed classifiers were evaluated on the *QUWI* and *MSHD* datasets. According to the experimental result, the proposed model that employed *SVM* obtained an accuracy of 68.75% for the *QUWI* dataset and an accuracy of 73.02% for the *MSHD* dataset when slant and curvature features were used.

Akbari et al. [28] introduced a gender classification methodology using handwriting images. This method initially transforms each image into a texture representation, which is then decomposed into multiple subbands at different levels. These subbands are subsequently utilized to create Probabilistic Finite State Automata (PFSA) for generating feature vectors. For classification purposes, they applied both *SVM* and NNs. The proposed models were trained and evaluated on the *QUWI* and *MSHD* datasets. According to the experimental result, the proposed NN obtained the best accuracy, an accuracy of 79.30% on the *QUWI* dataset. When it comes to the *MSHD* dataset, the proposed model based on *SVM* obtained the best accuracy, an accuracy of 79.90%.

Bouadjenek et al. [29] proposed a gender classification model for handwriting images. They employed *HOG* and *Local Binary Patterns* (*LBP*) as feature extractors and *SVM* as the classifier. The experiments were carried out on the *IAM* dataset. According to the experimental result, the proposed model based on *HOG* obtained an accuracy of 74%.

Youssef et al. [30] proposed a gender classification model based on the combination of *Wavelet Domain Local Binary Patterns*

(*WD-LBP*) and *SVM*. They trained and evaluated their model on a subset of the *QUWI* dataset, which consists of documents in English, and Arabic. According to the experimental result, the proposed model obtained an accuracy of 74.30%.

Illouz et al. [18] proposed a CNN-based model for gender classification using handwriting data. This model comprised six layers: Four convolutional layers, succeeded by a *Dense* layer and a *softmax* output layer. The proposed classifier underwent evaluation on their proprietary dataset, namely, *HEBIU*, encompassing a total of 810 samples in Hebrew and English collected from 405 subjects. Through a series of experiments, the highest accuracy achieved was 82.89%. Notably, this peak accuracy was attained when the model was trained on Hebrew samples and evaluated on English samples. Unlike our model, this model yielded 200 patches extracted from each handwriting sample.

Maken and *Gupta* [31] proposed an ensemble approach that employed *SVM*, *Logistic Regression* (*LR*), and *k-Nearest Neighbor* (*kNN*) for automated classification of gender from handwriting using the landmarks of differences between genders. They used the shape of the visual appearance of the handwriting for extracting features of the handwriting such as slanteness (direction), area, and perimeter. The proposed model was evaluated on the dataset of the *ICDAR 2013 Gender Prediction Competition*, which comprised 282 writers with 2 samples per writer. According to the experimental result, the proposed model obtained an accuracy of 65.71%.

Table 1 lists a comparison of the related work in terms of employed technique(s), used dataset(s), covered language(s), content type, and obtained gender classification accuracy.

Table 1 A comparison of the related work

Related Work	Employed Technique(s)	Used Dataset(s)	Covered Language(s)	Content Type	Classification Accuracy
[19]	Exponential Gradient	A subset of *BNC*	English	Handwritten text	~80%
[20]	*SVM*, and *GMM*	*IAM-OnDB*	English	Handwritten text	67.06%
[21]	*COLD* and Hinge features coupled with *SVM*	A subset of *QUWI*	English, and Arabic	Handwritten text	73.60%
[22]	CNN	*IAM*, and *KHATT*	English, and Arabic	Handwritten text	80.72% (*IAM*) 68.90% (*KHATT*)
[23]	CNN	*QUWI*, and a dataset of documents in Hebrew script	English, Arabic, and Hebrew	Handwritten text	N/A
[26]	*HOG*, and *SVM*	*IAM*, and *KHATT*	English, and Arabic	Handwritten text	75.45% (*IAM*) 68.89% (*KHATT*)
[27]	ANN, and *SVM*	*QUWI*, and *MSHD*	English, French, and Arabic	Handwritten text	68.75% (*QUWI*) 73.02% (*MSHD*)
[28]	PFSA, *SVM*, and NN	*QUWI*, and *MSHD*	English, French, and Arabic	Handwritten text	79.30% (*QUWI*) 79.90% (*MSHD*)
[30]	*WD-LBP*, and *SVM*	A subset of *QUWI*	English, and Arabic	Handwritten text	73.40%
[29]	*HOG*, *LBP*, and *SVM*	*IAM-OnDB*	English	Handwritten text	74%
[18]	CNN	*HEBIU*	Hebrew, and English	Handwritten text	82.89%
[31]	*SVM*, *LR*, and *kNN*	*ICDAR 2013*	English, and Arabic	Handwritten text	65.71%
[32]	CNN	*ICDAR 2013*, *IAM-OnDB*, and *KHATT*	English, and Arabic	Handwritten text	71.8% (*ICDAR 2013*), 76.1% (*IAM*), 74.1% (*KHATT*)
Proposed work	**Image preprocessing, and CNN**	**Own dataset, and *IAM-OnDB***	**Turkish, and English**	**Handwritten text and numeral**	**74.46% (proprietary dataset) 68.11% (*IAM-OnDB*)**

Xue et al. [32] proposed *ATP-DenseNet*, an attention-based two-pathway Densely Connected Convolutional Neural Network to identify the gender of handwriting. More specifically, they proposed three models based on this architecture as follows: (i) *ATP-DenseNet-121*, (ii) *ATP-DenseNet-169*, and (iii) *ATP-DenseNet-201*. The proposed models were evaluated on three widely used datasets, namely, (i) *ICDAR 2013*, (ii) *IAM*, and (iii) *KHATT*. According to the experimental results, *ATP-DenseNet-169* obtained accuracy scores of 71.8%, 76.1%, and 74.1% on the *ICDAR 2013*, *IAM-OnDB*, and *KHATT* datasets, respectively.

Unlike the related work, the proposed model (i) utilizes both handwritten text and numeral while the related work utilizes only handwritten text, (ii) was finalized through an extensive task of optimization task, and (iii) to the best of our knowledge, is the only study that covers Turkish.

3. Material and Method

In this section, we describe the details of (i) how the used dataset was constructed and which preprocessing techniques were employed, (ii) the software stack used for the implementation of the proposed models, (iii) the proposed novel models for gender classification, and (iv) the employed evaluation metrics to evaluate the performance of the proposed models.

3.1 Dataset Preparation

To the best of our knowledge, there does not exist a public handwritten Turkish text/numerals dataset that is labeled with the corresponding gender. Therefore, we constructed our own dataset through a designed form that was filled out by each volunteer. We have collected a total of 530 handwriting samples. Of these samples, 330 were male, 195 were female, and gender was not specified in 5 of them. Each form collected from volunteers is a single shape of paper that has a size of $A4$. While ordering the samples, a single type of research form was replicated from the same device was used. Each volunteer filled out the form with a single type of pen with a 0.9 mm $2B$ tip while sitting on a chair on a table. The parts that were required to be filled in handwriting in the form consist of two parts, namely, (i) text, and (ii) numbers, which were separated from each other by frames. The details regarding the information requested from the volunteers via the provided research form are listed in Table 2. Some sample handwritten text and numerals from the constructed dataset are presented in Figure 1, and Figure 2, respectively.

Table 2 The details regarding the information requested from the volunteers via the provided research form.

Information	Data Type	Values
Age range	Categorical	$15-17$, $18-21$, $22-29$, $30-50$, and $50+$
Educational status	Categorical	*primary education*, *secondary education*, *high school*, *associate degree*, *undergraduate*, and *graduate*

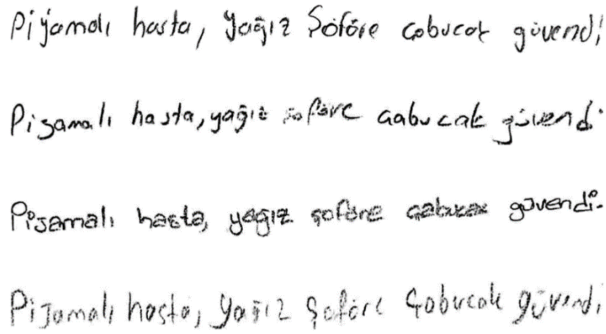

Figure 1 Some handwritten text samples from the constructed dataset

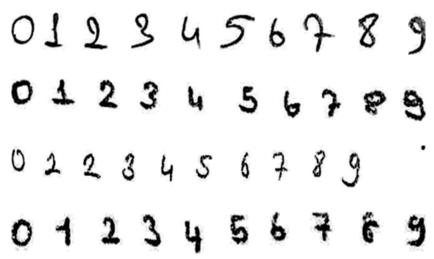

Figure 2 Some handwritten numeral samples from the constructed dataset

Additionally, volunteers were asked to fill in the text and numbers in the frame with their handwriting in a single line. The sample text used in this study consists of a sentence containing all 29 letters in Turkish. Regarding the numerical data to be filled, the volunteers were asked to write all the numbers in a single line. In the next step, each form was given an id number according to the collection order. These forms were scanned by 600 DPI *TA Triumph-Adler 4555i* printer, in color mode and PDF, at a high resolution of 1654 x 2338. Each page in the PDF has been converted to PNG format and named according to their id numbers and folders according to gender. To separate the text and numbers of each image in the folder, the coordinates were determined and cut by an implemented Python script that employed *pdf2image* [33], *OpenCV* [34], and *pandas* [35], [36] libraries to this end. The information in the form (such as age, and gender) was labeled with the id values and turned into categorical data. Then, the data were categorized and folded, and made ready for the preprocessing steps thanks to another implemented Python script. The process of constructing and preparing the novel dataset to be ready to be yielded into the proposed neural network is presented in Figure 3.

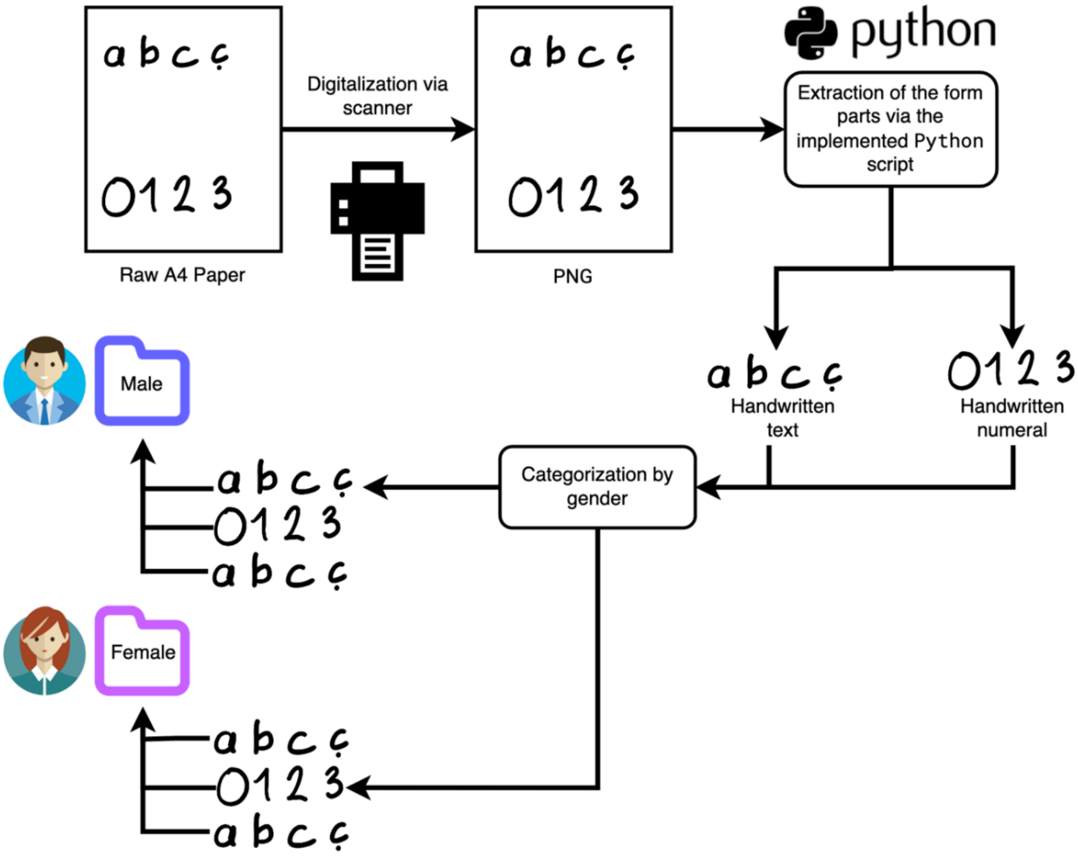

Figure 3 The process of constructing and preparing the novel dataset

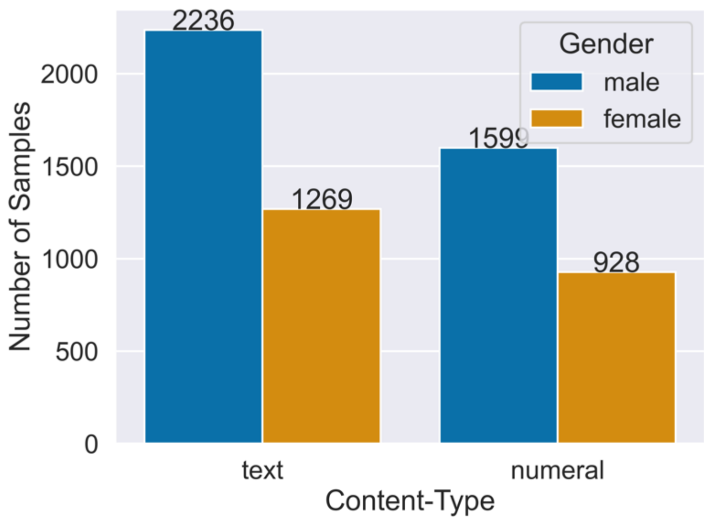

Figure 4 The distribution of collected handwriting samples by gender and content type

Similar to the approach by *Illouz et al.* [18], each handwriting sample was divided into square-shaped patches of $100x100$ pixels thanks to the implemented Python script as the widely-used pre-trained CNNs such as *ResNet50V2*, *InceptionV3*, and *MobileNetV2* do work with the square-shaped (*e.g.*, $32x32$, $75x75$, $299x299$) input, too. This operation also helped to keep the computational effort feasible. Then, each patch was mapped with the gender of the handwriting sample. Because of this process, the novel dataset was constructed. The distribution of the constructed dataset by gender and content type is presented in Figure 4.

Two rules were followed during the construction of the balanced dataset from the constructed dataset: (i) A handwritten text sample and a handwritten numerals sample were collected for each volunteer, and (ii) the same number of samples per gender were collected to construct a balanced dataset. Consequently, the constructed balanced dataset consisted of 928 handwritten text and 928 handwritten numerals per gender. The distribution of the constructed balanced dataset is given in Table 3.

Table 3 The distribution of the constructed balanced dataset

Gender	**Text**	**Numerals**	**Total**
$Male$	928	928	$1,856$
$Female$	928	928	$1,856$
Total	$1,856$	$1,856$	$3,712$

3.2 Software Stack

The entire software used for this study was implemented in the Python programming language and was powered by open-source technologies. *Keras* [37] was opted for the implementation of the proposed DNNs since being a high-level interface for the implementation of DNNs. Other advantages of *Keras* can be listed as follows: (i) seamless integration with other state-of-the-art data science frameworks, (ii) support to wide range of applications, including but not limited to computer vision, natural language processing, and plotting, and (iii) transferability as the models constructed using *Keras* can be easily transferred to various deep learning framework thanks to its modular design. The up-to-date version of *Keras* supports two deep learning backends, namely, (i) *TensorFlow* [38], and (ii) *Theano*. The selection of *TensorFlow* as *Keras*' backend arose from the developer's (*Keras*' creator) recommendation [39], owing to its ability to deliver high-performance and scalable capabilities. *NumPy* [40] and *pandas* [36], two widely-used Python libraries that *TensorFlow* depends on, were employed for the data manipulation and analysis of multi-dimensional matrices and numerical tables, respectively. In managing dataset operations such as partitioning into subsets based on the predefined ratio, data preprocessing tasks, and assessing the classification performance of the proposed models, a widely-used Python library, namely, *scikit-learn* [41] was employed. *Matplotlib* [42] was employed for the visualization of the experimental result. The details of the used software stack are listed in Table 4.

Table 4 The details of the used software stack

Software	**Version**
$Operating\ System$	$macOS\ Monterey$ 12.5
$Python$	3.8.13
$Keras$	2.8.0
$TensorFlow$	2.8.0
$NumPy$	1.21.5
$pandas$	1.4.2
$scikit-learn$	1.0.2

3.3 Proposed Model

We propose two novel CNN models for the gender prediction problem through the given handwriting samples: (i) The first model makes gender prediction through the given both handwritten text and numerals, and (ii) the second model makes gender prediction through the given handwritten text. Each proposed novel CNN model is described in the following subsections.

3.3.1 Proposed Two-Channel Model

The proposed two-channel model yields grayscale handwritten text and numerals in order to output its writer's gender into two classes, namely, (i) *male*, and (ii) *female*. To the best of our knowledge, this is the first gender prediction model that inputs handwritten numerals alongside handwritten text. The proposed two-channel model consisted of 21 layers as follows: Each channel is identical and consists of 9 layers. The model started with a *Convolutional* (denoted with *Conv2D*) layer, tasked with performing convolution operations on the provided input. This layer employed 16 filters and utilized a kernel size of $(3,3)$. Then, a *Batch Normalization* (denoted with *Batch Norm.*) layer was employed to normalize the activations of previous layers. Subsequently, a *Max Pooling* layer with a pool size of $(2,2)$ was applied to gradually diminish the spatial dimensions of the representation, ultimately leading to a reduction in the network's parameter count and computational workload [43]. Following this *Max Pooling layer*, a *Dropout* [44] layer with a dropout rate of 0.5 was employed to randomly drop neurons from the network, which eventually helps to prevent the well-known problem of DNNs, namely, the "overfitting". The second *Conv2D* layer with 32 filters and a kernel size of $(5, 5)$ succeeded the *Dropout* layer. Similar to the first Convolutional layer, a *Batch Norm.*, a *Max Pooling* with a pool size of $(2, 2)$, and a *Dropout* layer with a dropout rate of 0.5 succeeded the second *Conv2D* layer. As the final layer of each channel, a *Global Max Pooling* layer was utilized as the final layer, serving to consolidate activations across spatial locations and generate a vector of fixed size, which is common in several state-of-the-art CNNs [45]. A *Concatenation* layer was employed to concatenate the outputs of the channels. Another *Dropout* layer, but with a dropout rate of 0.6 followed the concatenation operation. Finally, a *Dense* layer, which is a deeply (fully) connected neural network component, with the sigmoid activation function was employed to output the predicted gender. The *Rectified Linear Unit* (*ReLU*) [46] was employed as the activation function of the *Conv2D* layers to avoid the vanishing gradient problem as a result of some other activation functions [47]. The default kernel initialization option of *Keras*, namely, *Glorot* (*a.k.a. Xavier*) *Uniform*, was employed as the kernel initializer of the employed *Conv2D* layers. Given that the handled problem is a binary classification task, the *Binary Cross-Entropy* was employed as the loss function of the model to calculate the loss after each epoch. The *Adadelta* [48] was employed as the optimization algorithm of the model to minimize the obtained loss by adjusting the attributes, namely, (i) weight, and (ii) bias. An overview of this model is presented in Figure 5. Table 5 lists each layer of the proposed two-channel model along with the hyperparameters that were utilized.

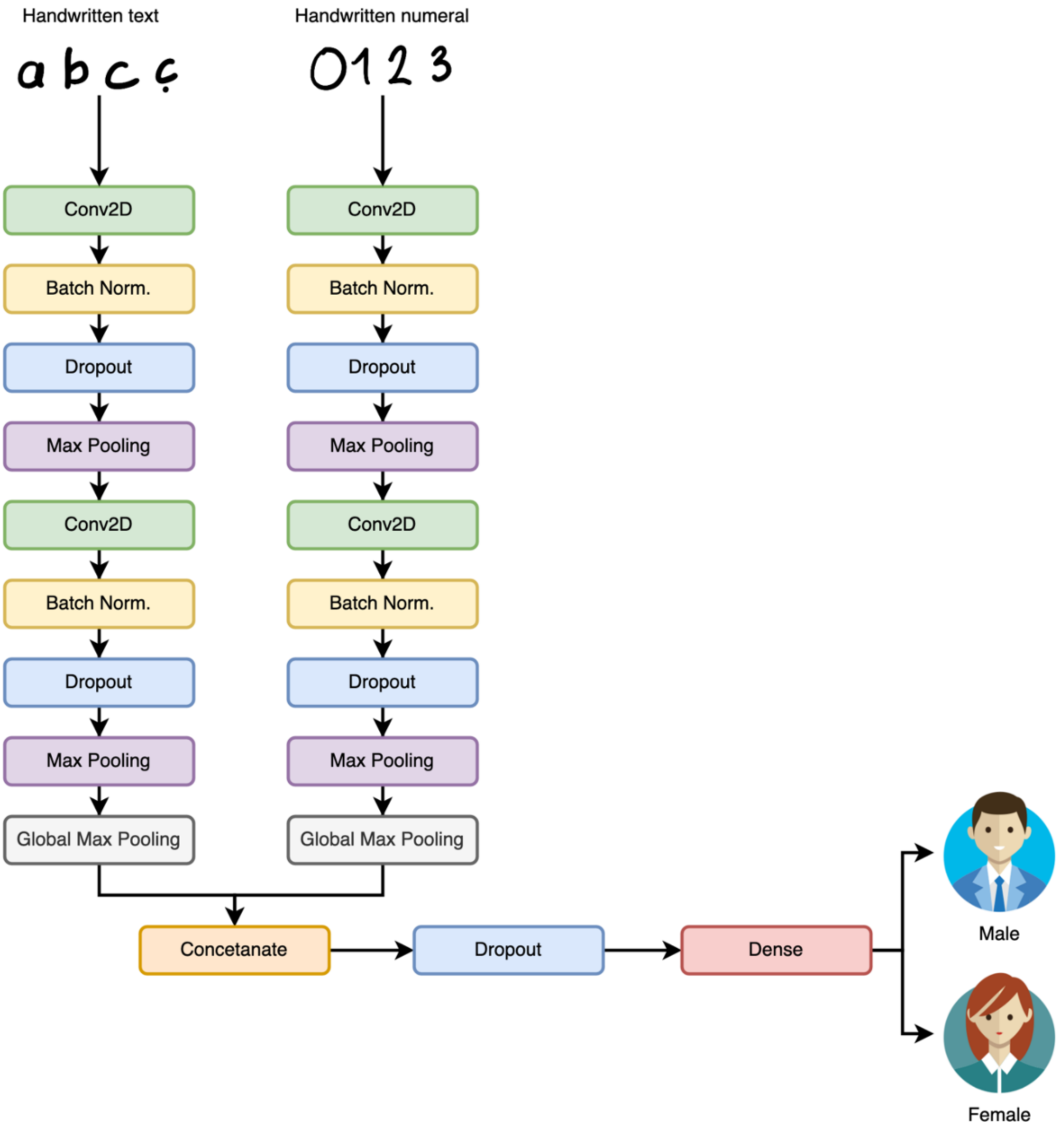

Figure 5 An overview of the proposed novel two-channel CNN model

Table 5 The layers of the proposed novel two-channel CNN model along with their corresponding hyperparameters

#	Layer Type	Hyperparameters
1	$Conv2D$	- $Number\ of\ filters$: 16 - $Kernel\ size$: $(2,2)$ - $Strides$: $(1,1)$ - $Kernel\ initializer$: $Glorot\ Uniform$ - $Padding$: $valid$ - $Activation\ function$: $ReLU$
2	$Batch\ Norm.$	N/A
3	$Dropout$	- $Dropout\ rate$: 0.6
4	$Max\ Pooling$	- $Pool\ size$: $(2,2)$
5	$Conv2D$	- $Number\ of\ filters$: 32 - $Kernel\ size$: $(2,2)$ - $Strides$: $(1,1)$ - $Kernel\ initializer$: $Glorot\ Uniform$ - $Padding$: $valid$ - $Activation\ function$: $ReLU$
6	$Batch\ Norm.$	N/A
7	$Dropout$	- $Dropout\ rate$: 0.6
8	$Max\ Pooling$	- $Pool\ size$: $(2,2)$
9	$Global\ Max\ Pooling$	N/A
10	$Conv2D$	- $Number\ of\ filters$: 16 - $Kernel\ size$: $(2,2)$ - $Strides$: $(1,1)$ - $Kernel\ initializer$: $Glorot\ Uniform$ - $Padding$: $valid$ - $Activation\ function$: $ReLU$
11	$Batch\ Norm.$	N/A
12	$Dropout$	- $Dropout\ rate$: 0.6
13	$Max\ Pooling$	- $Pool\ size$: $(2,2)$
14	$Conv2D$	- $Number\ of\ filters$: 32 - $Kernel\ size$: $(2,2)$ - $Strides$: $(1,1)$ - $Kernel\ initializer$: $Glorot\ Uniform$ - $Padding$: $valid$ - $Activation\ function$: $ReLU$
15	$Batch\ Norm.$	N/A
16	$Dropout$	- $Dropout\ rate$: 0.6
17	$Max\ Pooling$	- $Pool\ size$: $(2,2)$
18	$Global\ Max\ Pooling$	N/A
19	$Concatenate$	N/A
20	$Dropout$	- $Dropout\ rate$: 0.6
21	$Dense$	- $Number\ of\ units$: 1 - $Activation\ function$: $sigmoid$

3.3.2 Proposed Single-Channel Model

This model was intentionally proposed to be able to benchmark a CNN on a gold standard dataset as, to the best of our knowledge, there does not exist a handwritten gender dataset that contains both text and numeral. The proposed two-channel model consisted of 9 layers as follows: The model starts with a *Conv2D* layer with 32 filters and a kernel size of $(3,3)$. Then, a *Batch Norm.*, a *Dropout* layer with a dropout rate of 0.5, and a *Max Pooling* layer with a pool size of $(2,2)$ was employed, respectively. A second *Conv2D* layer followed this *Max Pooling* layer. Then, similar to the first Conv2D layer, a *Dropout* layer with a dropout rate of 0.5, and a *Max Pooling* layer with a pool size of $(2,2)$ was employed, respectively. Then, a *Global Max Pooling* layer was employed to aggregate the activations of spatial locations. Then, another *Dropout* layer with a dropout rate of 0.5 was employed. Finally, a *Dense* layer with a unit size of 1 and the *sigmoid* activation function was employed for the gender classification. Similar to the proposed two-channel model, (i) *ReLU* was employed as the activation function, (ii) the optimization of the model through the backpropagation was carried out on the *Adadelta* optimization algorithm, and (iii) the *Binary Cross-Entropy* was employed as the loss function of the model. An overview of this single-channel model is presented in Figure 6. Table 6 lists each layer of the proposed single-channel model along with the hyperparameters that were utilized.

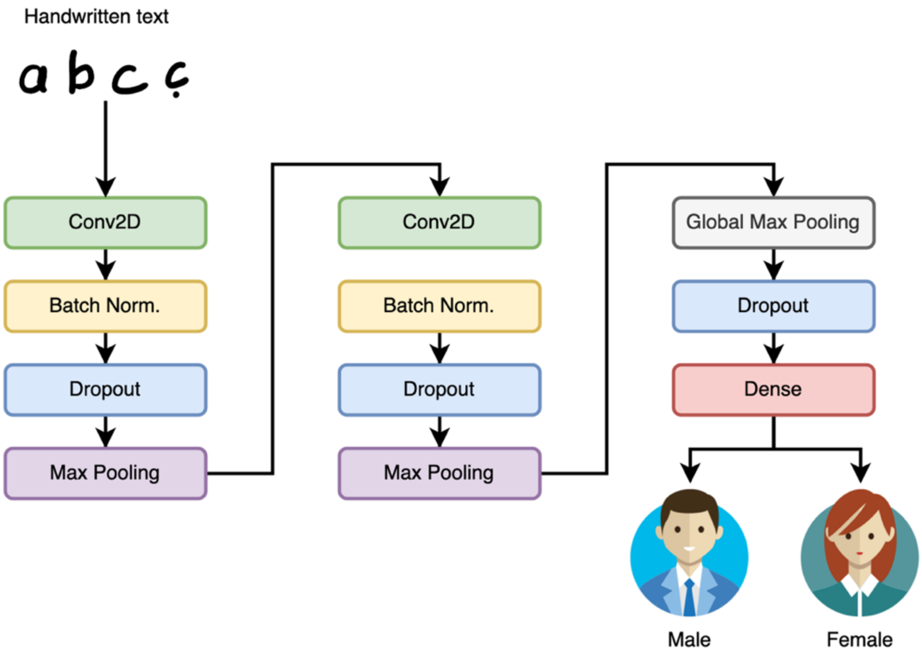

Figure 6 An overview of the proposed novel single-channel CNN model

Table 6 The layers of the proposed novel single-channel CNN model along with their corresponding hyperparameters

#	Layer Type	Hyperparameters
1	$Conv2D$	- $Number\ of\ filters$: 32 - $Kernel\ size$: $(2, 2)$ - $Strides$: $(1, 1)$ - $Kernel\ initializer$: $Glorot\ Uniform$ - $Padding$: $valid$ - $Activation\ function$: $ReLU$
2	$Batch\ Norm.$	N/A
3	$Dropout$	- $Dropout\ rate$: 0.3
4	$Max\ Pooling$	- $Pool\ size$: $(2, 2)$
5	$Conv2D$	- $Number\ of\ filters$: 64 - $Kernel\ size$: $(2, 2)$ - $Strides$: $(1, 1)$ - $Kernel\ initializer$: $Glorot\ Uniform$ - $Padding$: $valid$ - $Activation\ function$: $ReLU$
6	$Batch\ Norm.$	N/A
7	$Dropout$	- $Dropout\ rate$: 0.3
8	$Max\ Pooling$	- $Pool\ size$: $(2, 2)$
9	$Global\ Max\ Pooling$	N/A
10	$Dropout$	- $Dropout\ rate$: 0.3

3.4 Evaluation Metrics

$De-facto$ standard metrics to evaluate the performance of classifiers, namely, $accuracy$, $precision$, $recall$ (*a.k.a.* $sensitivity$), and $F1-score$ were employed to assess the classification performance of the proposed model. Let P denote *positives*, referring to the samples labeled with the target class, and N represent *negatives*, signifying the samples labeled with the complementary class of the target. TP, TN, FP, and FN denote correctly predicted *positives*, correctly predicted *negatives*, *positives* incorrectly predicted as *negative*, and *negatives* incorrectly predicted as *positive*, respectively. $Accuracy$ is the proportion of the correctly predicted samples to all samples. $Precision$ is defined as the proportion of accurately predicted positive instances to the total number of instances predicted as positive. $Recall$ is the ratio of correctly predicted positive instances to the total number of actual positive instances. $F1 - score$ is the harmonic mean of the $precision$ and $recall$ and is more useful than accuracy when the used dataset is imbalanced. The equations of $accuracy$, $precision$, $recall$, and $F1 - score$ are given in Eq. 1.

$$
\begin{aligned}
Accuracy &= \frac{TP + TN}{P + N} \\
Precision &= \frac{TP}{TP + FP} \\
Recall &= \frac{TP}{TP + FN} \\
F1 - score &= 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall}
\end{aligned} \tag{1}
$$

When it comes to the evaluation of the proposed model, we employed a confusion matrix, which is a specific table that visualizes the classification performance of classifiers. In the confusion matrix, every row corresponds to the count of samples in the true class, and each column corresponds to the count of samples in the predicted class.

4. Experimental Results and Discussion

In the following subsections, the hyperparameter optimization, the training and evaluation of the proposed models, and the discussion in the light of experimental results are described.

4.1 Hyperparameter Optimization

Hyperparameters are the parameters of a DNN model that impact the learning process and are determined through empirical tuning [39], [49]. More specifically, the hyperparameter optimization task provides various improvements to the neural network such as performance enhancement, generalization improvement, faster convergence, resource efficiency, robustness, and exploratory analysis. Both proposed models were trained under the same hyperparameters which were finalized as a result of automatized hyperparameter optimization task. Throughout this task, an extensive set of values for each hyperparameter was assessed to uncover the optimal combination of hyperparameters as they are listed in Table 7, where the obtained best value for each hyperparameter is given in bold. Despite encompassing a wide range of hyperparameters, the employed hyperparameter optimization task remained efficient and streamlined due to its automated nature. To be more specific, the proposed models utilized a widely-recognized technique known as $Hyperband$ [50] as the optimization algorithm. The optimization objective was defined as the accuracy achieved on the validation set. A subset of 20% from the training set was allocated for use as the validation set. As listed in Table 7, several widely-used optimization algorithms, namely, (i) $Adaptive\ Moment\ Estimation$ ($Adam$) [51], (ii) $Root\ Mean\ Square\ Propagation$ ($RMSprop$) [52], (iii) $Stochastic\ Gradient\ Descent$ (SGD) [53], and (iv) $Adadelta$ were evaluated as the optimization algorithms. The $Adadelta$ was employed as the optimization of the proposed model as a result of the employed hyperparameter optimization. Several widely-used activation functions, namely, (i) $Rectified\ Linear\ Unit$ ($ReLU$), (ii) $Exponential\ Linear\ Unit$ (eLU), (iii) $Parametric\ ReLU$ ($PReLU$), (iv) $Leaky\ ReLU$, (v) $tanh$, and (vi) $softmax$ were evaluated as the activation functions. The *ReLU* was employed as the activation algorithm of the proposed models as a result of the employed hyperparameter optimization. Regarding the $dropout\ rate$, an assessment encompassed a set of 0.2, 0.3, 0.4, 0.5, and 0.6. The range of 3 to 10 was scrutinized for the number of folds (k value). Additionally, evaluations were conducted for both $kernel\ regularization\ penalty$ and $bias\ regularization\ penalty$, considering the set of $1xe^{-5}$, $1xe^{-6}$, $1xe^{-7}$, and $1xe^{-8}$. The $batch\ size$ underwent evaluation using the set of 16, 32, 64, 128, and 256.

Table 7 The evaluated values of the employed hyperparameters during the optimization. The obtained best values were given in bold

Model	Evaluated Values
$Dropout\ rate$ for *Conv.* layers	0.2, 0.3, 0.4, 0.5, $\mathbf{0.6}$
$Dropout\ rate$ for the *Dense* layer prior to final	0.2, 0.3, 0.4, 0.5, $\mathbf{0.6}$
$Activation\ function$	$\boldsymbol{ReLU}$, eLU, $PReLU$, $Leaky\ ReLU$, $tanh$, $softmax$
$Optimization\ algorithm$	$Adam$, $RMSprop$, SGD, $\boldsymbol{Adadelta}$
$Kernel\ size$	$\mathbf{3}$, 5, 7, 9
$Kernel\ regularization\ penalty$	$1xe^{-5}$, $1xe^{-6}$, $\boldsymbol{1xe^{-7}}$, $1xe^{-8}$
$Bias\ regularization\ penalty$	$1xe^{-5}$, $1xe^{-6}$, $\boldsymbol{1xe^{-7}}$, $1xe^{-8}$
$Learning\ rate$	$\boldsymbol{1xe^{-3}}$, $5xe^{-3}$, $1xe^{-4}$, $1xe^{-5}$, $1xe^{-6}$
$Batch\ size$	16, 32, $\mathbf{64}$, 128, 256
$Number\ of\ folds$	3, 4, $\mathbf{5}$, 6, 7, 8, 9, 10

4.2 Model Training and Evaluation

The training setup of a neural network is crucial to ensure that the network is optimized for achieving its highest performance potential. Its proper design and configuration influence every aspect of neural network's performance, from accuracy and efficiency to generalization and robustness. A properly configured training can make the difference between a neural network that struggles to learn and one that excels at its intended task. A well-trained NN should neither overfit nor underfit. In the following subsections, the training and evaluation of the proposed models are described.

4.2.1 Two-Channel Model

The training of each proposed model was started with the *Early Stopping* callback, which is responsible for stopping the training when the model stops learning. In pursuit of this objective, two parameters are utilized in the following manner: (i) The monitored criterion, and (ii) the number of epochs that the callback waits before cessation, also known as "*patience*." Specifically, the achieved validation loss was designated as the monitored criterion, and a patience value of 10 epochs was established.

A subset comprising 20% of the entire dataset, totaling 372 samples, was set aside as the test set. This subset was employed to evaluate the classification performance of the proposed model. The remaining dataset was employed for both training and validation using the *Stratified k-Fold Cross Validation* technique, a specialized form of *k-Fold Cross Validation* that divides the entire dataset into k folds while maintaining the proportional distribution of samples for each class. The value of k was determined as 5 based on experimental results from the employed hyperparameter optimization. This indicates that the training set was divided into 5 folds, with the initial fold serving as the validation set and the remaining 4 folds being utilized for training purposes. This process was repeated 5 times to utilize the entire dataset for both training and validation purposes. Under this configuration, the training of the two-channel model was continued for 97 epochs until the employed *Early Stopping* callback stopped the training. An accuracy as high as 74.46% was obtained on the test set. The accuracy values achieved for both the training and validation sets during the training of the proposed two-channel model were plotted in Figure 7. According to this experimental result, it is safe to conclude that the model neither overfit nor underfit. The obtained confusion matrix for evaluating the test set is presented in Figure 8.

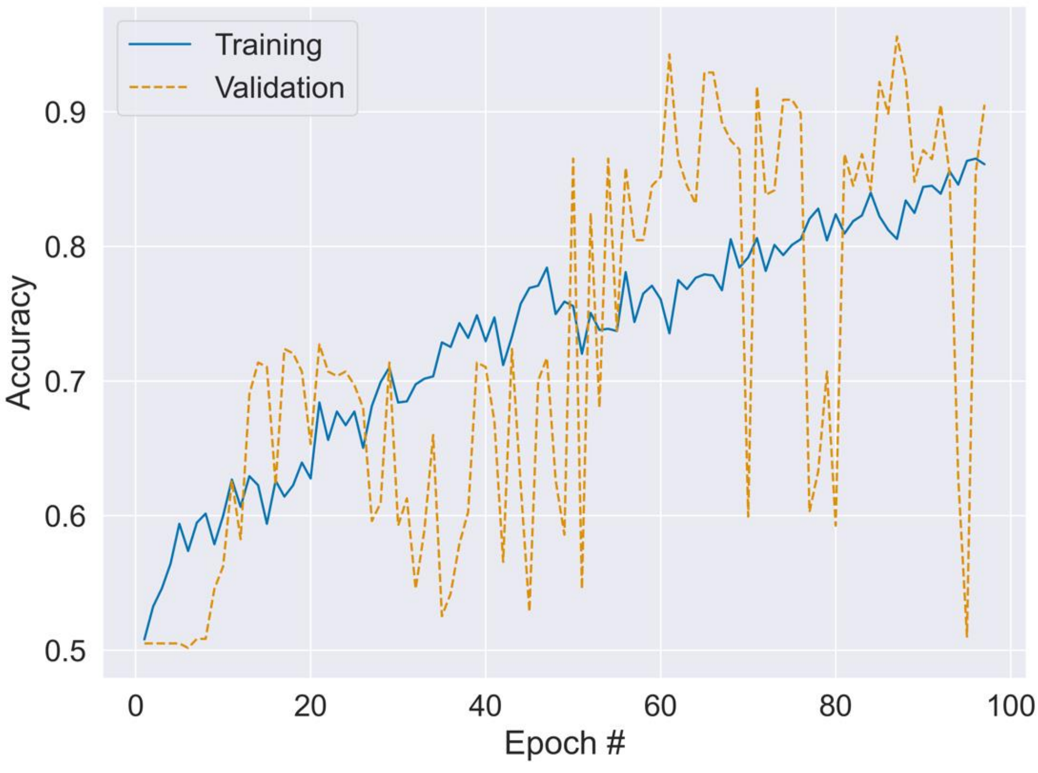

Figure 7 The accuracy values obtained for both the training and validation sets during the training process of the proposed two-channel model.

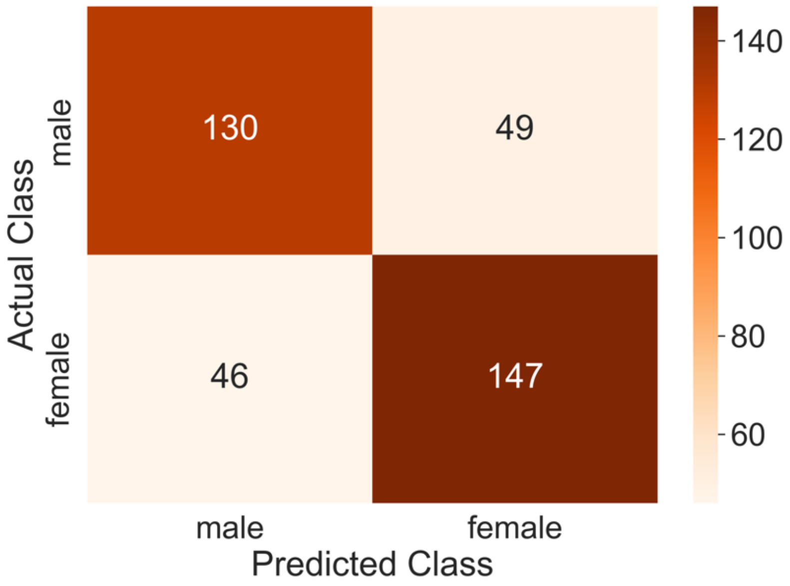

Figure 8 The obtained confusion matrix of the proposed two-channel model upon evaluation on the test set

4.2.2 Single-Channel Model

Under the same training configuration as the two-channel model, the single-channel model had been trained for 74 epochs until the employed *Early Stopping* callback stopped the training. An accuracy as high as 72.33% was obtained on the test set. The accuracy values achieved for both training and validation sets during the training of the proposed two-channel model were plotted in Figure 9. According to this experimental result, it is safe to conclude that the model neither overfit nor underfit. The obtained confusion matrix for evaluating the test set is presented in Figure 10.

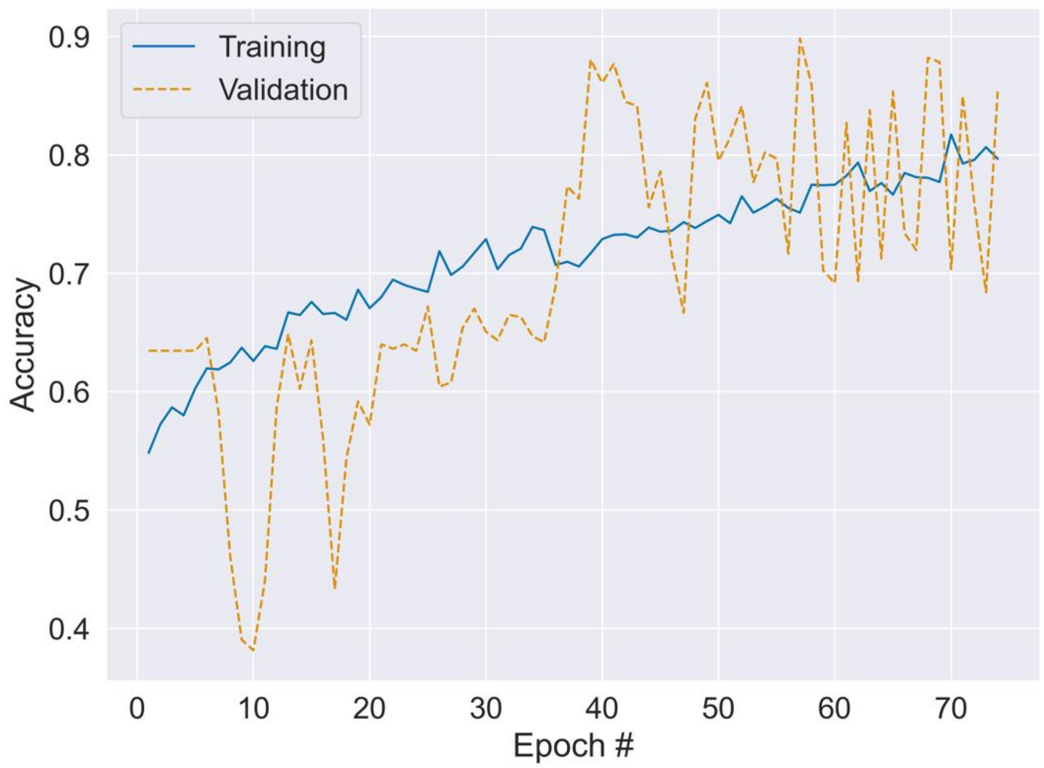

Figure 9 The accuracy values obtained for both the training and validation sets during the training process of the proposed single-channel model.

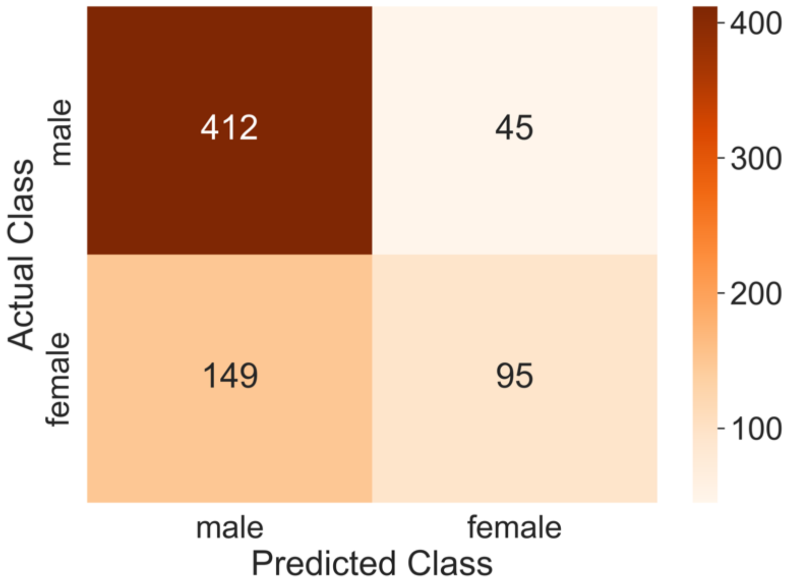

Figure 10 The obtained confusion matrix of the proposed single-channel model upon evaluation on the test set

We have also experimented with gender detection through the numerals only using the same single-channel model. This time, an accuracy of 64.03% was obtained. Alongside the proposed DNNs, (1) the widely-used traditional ML algorithms, namely, (i) *SVM*, (ii) *Logistic Regression*, (iii) *Naïve Bayes*, (iv) *Random Forest*, (v) *Decision Tree*, (vi) *k-Nearest Neighbors*, (vii) *Light Gradient Boosting Machine* (*LGBM*), and ($viii$) *eXtreme Gradient Boosting* (*XGBoost*), and (2) the widely-used pre-trained DNNs, namely, (i) *ResNet50V2* [54], (ii) *InceptionV3* [55], and (iii) *MobileNetV2* [56] through the transfer-learning were employed. For the pre-trained DNNs, each channel was replaced with the pre-trained DNN, including the weights calculated for the *ImageNet* dataset. The same layers of the proposed two-channel CNN were applied after the concatenation. Similarly, the pre-trained DNNs were trained under the same hyperparameters. As a final model, the proposed two-channel CNN was employed as the feature extractor, and *SVM* was employed as the classifier. According to the experimental result, the proposed two-channel CNN that yields both handwritten text and numerals provided the best accuracy among all models. The proposed single-channel model that yields handwritten text followed that. This experimental result demonstrates that yielding both handwritten text and numerals provides better accuracy than yielding only handwritten text or numerals. Another conclusion in the light of the experimental results is that the DNN models provided better accuracy than the traditional ML models for gender classification through handwriting. The obtained classification accuracy scores of the employed traditional ML algorithms and proposed CNN models on the test set of the novel dataset are listed in Table 8.

Table 8 The obtained classification accuracy scores of the employed traditional ML algorithms and proposed CNN models on the test set of the novel dataset

Model	**Accuracy**
SVM (Text)	53.49%
Logistic Regression (Text)	51.88%
Naïve Bayes (Text)	51.08%
Random Forest (Text)	58.33%
Decision Tree (Text)	52.42%
kNN (k=2*)* (Text)	52.96%
LGBM (Text)	55.11%
XGBoost (Text)	56.45%
ResNet50V2 (Text and Numeral)	47.58%
InceptionV3 (Text and Numeral)	50.54%
MobileNetV2 (Text and Numeral)	51.88%
Proposed Single-channel CNN (Text)	72.33%
Proposed Single-channel CNN (Numeral)	64.03%
Proposed Two-channel CNN (Text and Numeral)	**74.46%**
Proposed Two-channel CNN-SVM (kernel=*linear*) (Text and Numeral)	68.55%
Proposed Two-channel CNN-SVM (kernel=*rbf*) (Text and Numeral)	70.70%
Proposed Two-channel CNN-SVM (kernel=*poly*) (Text and Numeral)	71.51%

5. Conclusion

Gender is one of the most distinguishing characteristics of human beings. Drawing is the oldest communication tool of human beings and consists of a lot of information about its owner. From this point forth, we have proposed gender classification models through the given handwriting samples, which can be text, numerals, or both text and numerals. The proposed models were based on CNNs, which have provided the state-of-the-art for many classification problems whether it can be text classification or image classification. To the best of our knowledge, there does not exist a public Turkish handwriting dataset, which is labeled with genders. Therefore, we constructed our own dataset, which consists of 530 handwriting samples. The proposed CNN models were trained and evaluated on this dataset. According to the experimental result, the best accuracy, an accuracy as high as 74.46%, was obtained by the proposed CNN model that yields both handwriting text and numerals. The obtained accuracy is higher than the compared state-of-the-art techniques and is promising on such a task that even humans could not have achieved highly-accurate results for, as of yet. This experimental result demonstrates that yielding both handwriting text and numerals provides better accuracy compared to yielding only handwriting text or numerals. Another key finding in the light of the experimental result is that the models based on CNNs provided better accuracy compared to the models that employ the traditional ML algorithms as well as the combination of CNN and SVM.

As a future work, the authors would like to extend the constructed dataset by combining it with other datasets to improve the learning ability of the proposed model.

References

[1] J. Wayman, A. Jain, D. Maltoni, and D. Maio, *Biometric Systems*, 1st ed. London: Springer London, 2005. doi: 10.1007/b138151.

[2] R. N. King and D. J. Koehler, "Illusory correlations in graphological inference," *J Exp Psychol Appl*, vol. 6, no. 4, pp. 336–336, 2000, doi: 10.1037/1076-898X.6.4.336.

[3] V. Shackleton and S. Newell, "European Management Selection Methods: A Comparison of Five Countries," *International Journal of Selection and Assessment*, vol. 2, no. 2, pp. 91–102, 1994, doi: 10.1111/j.1468-2389.1994.tb00155.x.

[4] M. Ahmed, A. G. Rasool, H. Afzal, and I. Siddiqi, "Improving handwriting based gender classification using ensemble classifiers," *Expert Syst Appl*, vol. 85, pp. 158–168, 2017, doi: 10.1016/j.eswa.2017.05.033.

[5] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Local Descriptors to Improve Off-line Handwriting-based Gender Prediction," in *Proceedings of the 2014 6th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2014)*, Tunis, Tunisia: IEEE, pp. 43–47, 2014. doi: 10.1109/SOCPAR.2014.7007979.

[6] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Age, Gender and Handedness Prediction from Handwriting using Gradient Features," in *Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR '15)*, Washington, DC, United States: IEEE, 2015, pp. 1116–1120. doi: 10.1109/ICDAR.2015.7333934.

[7] "What is gender? What is sex?," *Canadian Institutes of Health Research*, 2020 [Online]. Available: https://cihr-irsc.gc.ca/e/48642.html. [Accessed: 10-Aug-2022].

[8] L. A. M. Galea and D. Kimura, "Sex differences in route-learning," *Pers Individ Dif*, vol. 14, no. 1, pp. 53–65, 1993, doi: 10.1016/0191-8869(93)90174-2.

[9] E. Coluccia, G. Iosue, and M. Antonella Brandimonte, "The relationship between map drawing and spatial orientation abilities: A study of gender differences," *J Environ Psychol*, vol. 27, no. 2, pp. 135–144, 2007, doi: 10.1016/j.jenvp.2006.12.005.

[10] R. S. Astur, A. J. Purton, M. J. Zaniewski, J. Cimadevilla, and E. J. Markus, "Human sex differences in solving a virtual navigation problem," *Behavioural Brain Research*, vol. 308, pp. 236–243, 2016, doi: 10.1016/j.bbr.2016.04.037.

[11] D. A. Cahn-Weiner, K. Williams, J. Grace, G. Tremont, H. Westervelt, and R. A. Stern, "Discrimination of Dementia with Lewy bodies from Alzheimer disease and Parkinson disease using the Clock Drawing Test," *Cognitive and Behavioral Neurology*, vol. 16, no. 2, pp. 85–92, 2003, doi: 10.1097/00146965-200306000-00001.

[12] K. Amunts *et al.*, "Gender-specific left-right asymmetries in human visual cortex," *Journal of Neuroscience*, vol. 27, no. 6, pp. 1356–1364, 2007, doi: 10.1523/JNEUROSCI.4753-06.2007.

[13] A. C. Hurlbert and Y. Ling, "Biological components of sex differences in color preference," *Current Biology*, vol. 17, no. 16, pp. 623–625, 2007, doi: 10.1016/j.cub.2007.06.022.

[14] S. Bradley, "Handwriting and Gender: A multi-use data set," *Journal of Statistics Education*, vol. 23, no. 1, pp. 1–15, 2015, doi: 10.1080/10691898.2015.11889721.

[15] G. Cordasco, M. Buonanno, M. Faundez-Zanuy, M. T. Riviello, L. Likforman-Sulem, and A. Esposito, "Gender Identification through Handwriting: an Online Approach," in *Proceedings of the 11th IEEE International Conference on Cognitive Infocommunications (CogInfoCom 2020)*, Mariehamn, Finland: IEEE, pp. 197–202, 2020. doi: 10.1109/CogInfoCom50765.2020.9237863.

[16] S. Saha, M. Asif, B. Khaled, M. S. Islam, N. Saha Puja, and M. Hasan, "Detecting Sex From Handwritten

Examples; Detecting Sex From Handwritten Examples," in *2018 IEEE International Conference on System, Computation, Automation and Networking (ICSCAN)*, Pondicherry, India: IEEE, pp. 1–7, 2018.
[17] S. Upadhyay, J. Singh, and S. K. Shukla, "Determination of Sex Through Handwriting Characteristics," *Int J Curr Res Rev*, vol. 9, no. 13, pp. 11–18, 2017, doi: 10.7324/ijcrr.2017.9133.
[18] E. Illouz, E. Omid David, and N. S. Netanyahu, "Handwriting-Based Gender Classification Using End-to-End Deep Neural Networks," in *Proceedings of the 27th International Conference on Artificial Neural Networks (ICANN 2018)*, Island of Rhodes, Greece: Springer Verlag, pp. 613–621, 2018. doi: 10.1007/978-3-030-01424-7_60.
[19] M. Koppel, S. Argamon, and A. R. Shimoni, "Automatically Categorizing Written Texts by Author Gender," *Literary and Linguistic Computing*, vol. 17, no. 4, pp. 1–13, 2002, doi: 10.1093/llc/17.4.401.
[20] M. Liwicki, A. Schlapbach, P. Loretan, and H. Bunke, "Automatic Detection of Gender and Handedness from On-Line Handwriting," in *Proceedings of the 13th Biennial Conference of the International Graphonomics Society (IGS2007)*, Melbourne, Australia, pp. 179–183, 2007.
[21] A. Gattal, C. Djeddi, A. Bensefia, and A. Ennaji, "Handwriting Based Gender Classification Using COLD and Hinge Features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Springer, pp. 233–242, 2020. doi: 10.1007/978-3-030-51935-3_25.
[22] Á. Morera, Á. Sánchez, J. F. Vélez, and A. B. Moreno, "Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks," *Complexity*, vol. 2018, pp. 1–14, 2018, doi: 10.1155/2018/3891624.
[23] I. Rabaev, M. Litvak, S. Asulin, and O. H. Tabibi, "Automatic Gender Classification from Handwritten Images: A Case Study," in *Proceedings of the 19th International Conference on Computer Analysis of Images and Patterns (CAIP 2021)*, Virtual, 2021, pp. 329–339, 2021. doi: 10.1007/978-3-030-89131-2_30.
[24] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proceeding of the 2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, Miami, FL, USA: IEEE, pp. 248–255, 2009. doi: 10.1109/cvpr.2009.5206848.
[25] S. Al Maadeed and A. Hassaine, "Automatic prediction of age, gender, and nationality in offline handwriting," *EURASIP J Image Video Process*, vol. 2014, no. 10, pp. 1–10, 2014, doi: 10.1186/1687-5281-2014-10.
[26] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Histogram of Oriented Gradients for Writer's Gender, Handedness and Age prediction," in *Proceedings of the 2015 International Symposium on Innovations in Intelligent SysTems and Applications, Proceedings (INISTA 2015)*, pp. 1–5, 2015. doi: 10.1109/INISTA.2015.7276752.
[27] I. Siddiqi, C. Djeddi, A. Raza, and L. Souici-meslati, "Automatic analysis of handwriting for gender classification," *Pattern Analysis and Applications*, vol. 18, pp. 887–899, 2015, doi: 10.1007/s10044-014-0371-0.
[28] Y. Akbari, K. Nouri, J. Sadri, C. Djeddi, and I. Siddiqi, "Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata," *Image Vis Comput*, vol. 59, no. C, pp. 17–30, 2017, doi: 10.1016/j.imavis.2016.11.017.
[29] N. Bouadjenek, H. Nemmour, and Y. Chibani, "Writer's Gender Classification Using HOG and LBP Features," in *Proceedings of the International Conference on Electrical Engineering and Control Applications (ICEECA 2016)*, Kuala Lumpur, Malaysia, pp. 1–5, 2016. doi: 10.1007/978-3-319-48929-2_24.
[30] A. E. Youssef, A. S. Ibrahim, and A. Lynn Abbott, "Automated Gender Identification for Arabic and English Handwriting," in *Proceedings of the 5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013)*, London, UK, pp. 1–6. Doi, 2013: 10.1049/ic.2013.0274.
[31] P. Maken and A. Gupta, "A method for automatic classification of gender based on text- independent handwriting," *Multimed Tools Appl*, vol. 80, no. 16, pp. 24573–24602, Jul. 2021, doi: 10.1007/s11042-021-10837-9.
[32] G. Xue, S. Liu, D. Gong, and Y. Ma, "ATP-DenseNet: a hybrid deep learning-based gender identification of handwriting," *Neural Comput Appl*, vol. 33, pp. 4611–4622, May 2021, doi: 10.1007/s00521-020-05237-3.
[33] E. Belval, "pdf2image: A python (3.6+) module that wraps pdftoppm and pdftocairo to convert PDF to a PIL Image object," 2021. [Online]. Available: https://github.com/Belval/pdf2image [Accessed: 10-Aug-2022].
[34] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, vol. 120, pp. 122–125, 2000.
[35] The pandas development team, "pandas: Python Data Analysis Library," 2020. [Online]. Available: https://pandas.pydata.org. [Accessed: 12-Jul-2022].
[36] W. McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference (SCIPY 2010)*, Austin, Texas, pp. 56–61, 2010. doi: 10.25080/majora-92bf1922-00a.
[37] F. Chollet, "Keras: the Python deep learning API," 2015. [Online]. Available: https://keras. [Accessed: 2-Dec-2022].
[38] M. Abadi *et al.*, "TensorFlow: A System for Large-Scale Machine Learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 2016)*, Savannah, GA, USA, pp. 265–283, 2016.
[39] F. Chollet, *Deep Learning with Python*. Manning Publications, 2017.
[40] C. R. Harris *et al.*, "Array Programming with NumPy," *Nature*, vol. 585, pp. 357–362, 2020, doi: 10.1038/s41586-020-2649-2.
[41] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12,

pp. 2825–2830, 2011.
[42] J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” *Comput Sci Eng*, vol. 9, no. 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.
[43] “CS231n Convolutional Neural Networks for Visual Recognition,” *Stanford University*, 2020. https://cs231n.github.io/convolutional-networks (accessed Dec. 02, 2022).
[44] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting,” *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.
[45] V. Christlein, L. Spranger, M. Seuret, A. Nicolaou, P. Král, and A. Maier, “Deep Generalized Max Pooling,” in *Proceedings of the 15th International Conference on Document Analysis and Recognition (ICDAR 2019)*, Sydney, Australia, pp. 1–7, 2019.
[46] V. Nair and G. E. Hinton, “Rectified Linear Units Improve Restricted Boltzmann Machines,” in *Proceedings, of the 27th International Conference on Machine Learning (ICML 2010)*, Madison, WI, USA, pp. 807–814, 2010.
[47] A. Mollahosseini, D. Chan, and M. H. Mahoor, “Going deeper in facial expression recognition using deep neural networks,” in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV 2016)*, Lake Placid, NY, USA: IEEE, pp. 1–10, 2016. doi: 10.1109/WACV.2016.7477450.
[48] D. C. Johny *et al.*, “ADADELTA: An Adaptive Learning Rate Method,” *IEEE Access*, vol. 7, no. November, 2018.
[49] X. Zhang, X. Chen, L. Yao, C. Ge, and M. Dong, “Deep Neural Network Hyperparameter Optimization with Orthogonal Array Tuning,” in *International Conference on Neural Information Processing (ICONIP 2019)*, Sydney, NSW, Australia: Springer, pp. 287–295, 2019. doi: 10.1007/978-3-030-36808-1_31.
[50] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, “Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization,” *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6765–6816, 2018.
[51] D. P. Kingma and J. L. Ba, “Adam: A Method for Stochastic Optimization,” in *Proceedings of the 3rd International Conference on Learning Representations (ICLR 2015)*, San Diego, California, USA, pp. 1–15, 2015.
[52] G. Hinton, N. Srivastava, and K. Swersky, “Neural Networks for Machine Learning,” 2012.
[53] H. Robbins and S. Monro, “A Stochastic Approximation Method,” *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951, doi: 10.1214/aoms/1177729586.
[54] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks,” in *Proceedings of the 14th European Conference on Computer Vision (ECCV 2016)*, Amsterdam, The Netherlands, Oct. 2016. doi: 10.1007/978-3-319-46493-0_38.
[55] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, Seattle, Washington, USA: IEEE, pp. 2818–2826,2016. doi: 10.1109/CVPR.2016.308.
[56] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, “MobileNetV2: Inverted Residuals and Linear Bottlenecks,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2018)*, Salt Lake City, UT, USA: IEEE, pp. 4510–4520, 2018. doi: 10.1109/CVPR.2018.00474.

Acknowledgments
The authors would like to thank a large number of anonymous subjects from Duzce University who voluntarily participated in this study by providing the handwriting samples.

Conflict of Interest Notice
The authors declare that there is no conflict of interest regarding the publication of this paper.

Ethical Approval and Informed Consent
It is declared that during the preparation process of this study, scientific and ethical principles were followed, and all the studies benefited from are stated in the bibliography.

Availability of data and material
Not applicable

Plagiarism Statement
This article has been scanned by iThenticate ™.