International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Gender Classification Based on Online Signature Features using Machine Learning Techniques

Shivanand S. Gornale[1], Sathish Kumar*[2], Rashmi Siddalingappa[+2], Arjun Mane[3]

Submitted: 05/05/2022 Accepted : 12/05/2022

Abstract: A human signature gives a lot of insights into an individual's characteristics such as illness, professional choices, and mood. From the biometric perspective, a Handwritten Signature is a behavioral trait and Gender is a demographic category (soft trait) of the person. Gender classification from handwritten signatures has been implied in several applications such as psychology and forensics. Male writings with a high intra-class variation tend to have a feminist aesthetic aspect, and vice versa. This gives clues to recognize the gender of the person using a handwritten signature. The proposed methodology is based on extracting numeric features from the male and female dynamic signature samples. Data was collected from 535 individuals of different age groups (18-65). Further, these signature samples were converted to numeric attributes resulting in 66 signature features from each data. Experiments were carried out using six different Machine Learning techniques; On the whole, the overall accuracy of these methods is 81.2% (KNN), 81.9% (LR), 77.1% and 49.3% (for both Poly and RBF kernels in SVM, respectively), Poly kernel using cross-validation resulted in 81.8% in SVM, 89.3% (DT), 96.2% (RF) and 98.2% (DL). Overall, the deep neural networks outperformed other models, immediately followed by RF.

Keywords: *Biometric Data Analysis, Gender Classification, Online Handwritten Signature, Feed Forward Deep Neural Network*

1. Introduction

A human signature gives a lot of insights into an individual's characteristics such as illness, professional choices, mood, and gender. This area is termed graphology. Studies related to graphology are limited around the world. Further, there are no institutions that provide education on graphology because these corporations do not benefit from graphology at all. Despite these challenges, the research on signatures includes often overlooked information.

Over the past decade, Research in Biometric security systems and related applications have worked to exploit the individualism of human beings. Biometrics are the biological characteristics used for recognizing and authentication based upon physiological traits Iris, Face, Fingerprint, hand geometry and behavioural traits Gait, Voice, keystroke dynamics and Signature of an individual [1]. In recent years, Human behaviour and social interventions of the individuals in a group are also identified under surveillance systems. Biometrics is used for recognition and reveals other demographic attributes like gender, age, nationality, and mental state of the person. Psychological, environmental and physical, and other factors are involved in the careful selection of biometrics, used in many specific applications [2]. Among these, handwritten signatures are a mature behavioural biometric modality, which forms a particular case of handwriting, which results in activation of brain regions that are conceived and controlled the masculine movements [37]. Everyone has an awareness of the gesture developed in the brain, which makes the writing of the individual unique [3]. Over the years, multiple research studies were carried out on a person's demographic identification using their handwriting. The acquisitions of handwritten signatures are typically in two significant categories: Static/offline and Dynamic/online signatures. Offline handwritten signatures are captured from the scanning manuscripts procedure in the static case. In an online handwritten signature, the characteristics are captured by setting some parameters (X and Y coordinates, Speed, Pressure, Time series, Azimuth, Pen Up & down) in the signatures using specialized acquisition devices. However, due to significant improvements in the acquisition technology, on several occasions, the usage of dynamic systems in forensic applications and many other domains it has become inevitable [4].

The online handwritten signature contains inherent behavioural features of the person, which can be used to analyze soft characters such as gender, age, handedness, personality analysis, etc. [5], [6]. Gender determination is an essential requirement in the analysis of forensic documents, such as suicide notes, sinister letters and malicious handwritten messages, property or tenancy, and more, where it is necessary to identify the person [7]. In many cases, shreds of evidence were created to point toward the innocent of misguiding the crime investigation. Even in personal financial cases, forged handwritten signatures are executed on property documents and multiple cheques.

From the studies, it is observed that men and women have popular stereotypes of handwriting skills. Women's signature is more attractive, neat, uniform, and legible in form than men's signature. In all such cases, a comparison of handwritten signatures identifies the person's characteristics. Hence, in the present study gender

[1] *Department of Computer Science, Rani Channamma University, Belagavi-591156, India, ORCID ID:0000-0001-5373-4049*
[2] *Department of Computer Science, Rani Channamma University, Belagavi-591156, India, ORCID ID:0000-0001-9374-1980*
[3] *Department of Computational and Data Sciences, Indian Institute of Science, Bengaluru 560012, India, ORCID ID:0000-0001-9786-8436*
[4] *Department of Digital and Cyber Forensics, Govt. Institute of Forensic Science, Nagpur-India, ORCID ID:0000-0003-4129-1863*
* *Corresponding Author Email: sathishkumarst25@gmail.com*

classification using different machine learning techniques from online handwritten signatures has been proposed.

The rest of the paper is organized into following sections. Section 2 concentrates on the literature work reviewing the studies on gender classification using handwritten signatures. The proposed model and the related methodology are reported in Section 3. Section 4 delineates the proposed experimental protocol and results achieved. The final section concludes the research work and talks about the scope for future work.

2. Related Work

More recently, some researchers have initially proven the system's ability by studying the challenges in online handwritten signatures for gender classification through the layout of contemporary architectures.

In [8], the authors proposed work on predicting person's emotional states using online handwriting and handwritten signature. The database comprised 804 handwriting and 8040 handwritten signatures, including demographic information of the person. K-NN, Jrip, and Random Forest classifiers are used for classifications. An accuracy of 55-58% for handwriting and 45-50% for signatures are achieved.

In [9], the sequential forward selection algorithm MRMR (Minimum Redundancy Maximum Relevancy) is used for Age and Gender-based handwritten signature systems using invariant features. TCDatabase consists of 50 users with every 40 signatures of different gender and age group. The features are classified using a global classifier; an error rate of 6.25% for gender and 2.6% for age classification is obtained.

In [10], authors have proposed work on online multilingual writer identification using a Deep Neural Network. Fuzzy-based elementary perceptual codes and Extended Beta-Elliptic model feature extraction techniques were used on two online handwriting standard databases IBM_UB_1 and ADAB. An accuracy of 96.90% for the IBM_UB_1 database and 98.25% for the ADAB dataset is achieved.

In [11], authors have proposed work on online handwriting-based automatic detection of gender and handedness of person. IAM-OnDB, a large online handwriting database, is used for the experimentation. 29 different features are extracted and classified using two techniques Support Vector Machine and Gaussian Mixture Model. GMM model gives the highest accuracy, achieving 67.06% for gender classification and 84.66 % for handedness classification.

Handwritten signatures may vary with time progression because of aging while considering a specific person. There are certain databases based on online handwriting and less number of handwritten signature datasets with gender annotations from the related work. Few researchers are working on online signature verification systems [12-16]. By observing previous works, Gender classification using online handwritten signatures appears to have received significantly less attention. The present study aims to develop a framework for detecting the handwritten signatures of male and female signatures of varying ages. The contribution of this proposed work is to (i) Develop a machine learning based framework that uses multi-features to analyze a person's gender classification based on their handwritten signatures. (ii)The comparative study of Logistic Regression, Decision Tree, k-Nearest Neighbor, Support Vector Machine, Random Forest, and Feed-Forward Deep Neural Network techniques are used for classification. (iii) Creation of a real-time dataset with different demographic annotations

3. Materials and Methods

i. Signature and Signature attributes and its relevance in Gender Classification

Human interaction with systems became more popular in day-to-day life due to rapid growth in pattern recognition and machine learning technologies. An Individual's handwritten signature is a distinctive and measurable behavioural biometric trait used to describe the person's behaviour and health conditions. Signatures are into two type static/offline and online/dynamic signatures. The static data is acquired from the scanned document image. Online signature data exploits rich information while writing on the digitized signature device. Much information such as X-Y coordinates, form, slant, spacing, azimuth, stroke, speed, pressure, angle, etc., is recorded. Compared to offline signature data, it usually has more accuracy. Gender classification from signature data analysis is an effective and efficient strategy in biometrics. The purpose of the gender classification from handwritten signatures is to check the resemblance between the male and female signatures. [17].

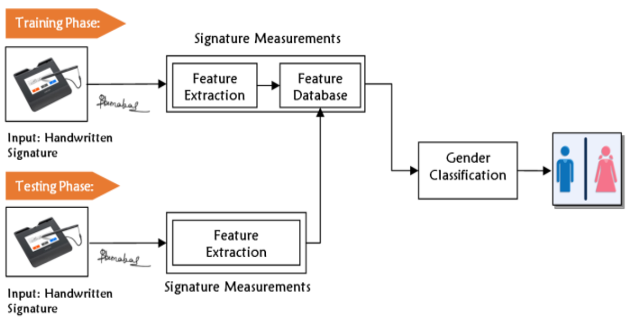
ii. Machine Learning and its relevance in Gender Classification

Human subjects have been widely using automated computer-driven image recognition systems in recent years. Many algorithmic systems are used to aid in user identification and authentication. Machine Learning (ML) techniques are the sub-field of Artificial Intelligence. These models are often trained to perform the tasks with many examples of pictures, objects, and scenarios. The trained data provides them with information about identifying common objects, but it lacks the necessary information for classifying genders. A specific type of machine learning technique (supervised learning, unsupervised learning, reinforcement learning) used by data scientists typically begins with constructing a model by selecting a dataset that depicts classification performed correctly. The main goal is to create a general rule related to inputs and outputs [18]. Machine Learning models play a major role in improving the performance of biometric systems. Some monotonous tasks such as one-to-one verification and one-to-many identification can be done automatically with embedded ML in biometrics. Much large dataset is needed to train and test a good ML model. Many authors have proposed efficient frameworks for ML-based gender classification algorithms [19-20].

3.1. Proposed Methodology

Figure 1 depicts the block diagram of steps involved in the proposed work.

Fig. 1. Proposed architecture model for gender classification



3.2. Dataset Description

The related work observed that a few standard datasets are available with minimum samples with limited demographic notations. With this motivation, the own signature dataset used in the present research work is collected and processed based on the different orientation features of an individual's signature.

- The signature samples are collected from persons of different age groups, and the samples include multilingual scripts such as Kannada, Hindi, Marathi, and English.
- All individuals who had knowledge of how to write a signature on digital device and were educated on the purpose of collecting signature samples.
- All the participants' consent form is obtained to make sure that professional codes of conduct and ethical principles are accepted and implemented.
- 10 Signature samples were recorded on Signature Device (Wacom STU-540) from each individual. Gender and age were also recorded.
- The signature was collected from 535 volunteers, including both males and females (282 males and 253 females).
- Along with signature samples from each volunteer, ten different fingerprint Samples using the Secugen-Hamster Fingerprint device are collected.
- Ten facial image samples of each volunteer are also collected using a Cannon 1300D DSLR camera at different distances and angles.
- From each sample, Blood pressure (both systolic and diastolic), and pulse rate are also collected using USB-based Apollo Pharmacy's Blood Pressure Monitor device.

62 system-generated features were collected in a .CSV file from each signature sample. Here, we have focused on the statistical features of the signature. The four different demographic features such as gender, age, Blood Pressure values, and pulse rate are recorded in the same .CSV file. Therefore, a total of 66 features are contained in .CSV file from each individual [36]. For the implementation of the present study, gender has been marked as labels, and other fields are considered as training features.

3.3. Feature Extraction

The feature extraction technique is integral to the gender classification approach [1]. This includes reducing the number of resources required to describe large data sets. Table 1. Shows the set of features extracted in online handwritten signature data.

3.4. Implementation platforms:

The entire model was executed on the anaconda navigator, a GUI distribution to support many implementation platforms (https://docs.anaconda.com/anaconda/navigator/install/). Spyder 3.9 is used for the interactive execution environment (https://www.spyder-ide.org/). The code is written in python language (https://www.python.org/). The required libraries such as sklearn, pandas, Numpy, and matplotlib were used in each algorithm using the import statement.

3.5. Classification:

3.5.1 K-Nearest Neighbor (k-NN) Classifier:

The k-NN algorithm is used to classify the data points in a dimensional space based on a different distance between one point and another on selected purpose [1] [21]. In this work, k-NN will classify by suitable K-value, which finds the nearest neighbour class label un-labeled. Empirically, City-Block distance K=8 is considered for the experiment. k-NN classifies feature vectors based on similarity measures. k-NN is a non-parametric algorithm that uses the following equation1 to calculate the shortest d distance between the j^{th} instance of the training sample M and the j^{th} instance of the testing pattern N, where 'n' is the total number of samples in the dataset:

$$D_{City-Block}(M,N) = \sum_{J=0}^{n} |M_J - N_J| \qquad (1)$$

3.5.2 Decision Tree:

The Decision tree is one of the renowned and widely used supervised learning for binary-class problems. The decision tree deduces the observations against the conclusion of the targeted value. DT has recursively split a dataset based on the criteria until it is left with leaf nodes. It grows with training data and starts at root node [22]. Gini Index and Entropy Index are commonly used criteria in most of the parts:

$$Gini\ Index: G(E) = 1 - \sum_{j=1}^{c} p_j^2 \qquad (2)$$

$$Entropy\ Index: H(E) = -\sum_{j=1}^{c} p_j log p_j \qquad (3)$$

Where, E is the particular instance/sample requiring the creation of a node, c is current state, and P_j is the Probability of an event j of state c or Percentage of class j in a node of state c. The Decision tree creates a root on randomly selected n training data from the feature dataset and assigns each node with sample data. Repeat until each node has a single sample from the same class. It chooses likely variables at random and selects the best split feature samples using the Gini Index and the Entropy index, which splits into two child nodes by passing the corresponding subsets.

3.5.3 Support Vector Machine:

The SVM is a binary classification method, which works on a multi-dimensional feature vector's decision boundary with classes in training set and known class labels. The SVM algorithm separates the closest training patterns with the highest margin using hyper-plane [23, 24].

3.5.4 Random Forest:

Random Forest (RF) algorithm conglomerates bagging of tree-structured classifiers. It has an impressive prediction model performance; it can handle many training samples and high dimensional spaces [7]. While building the decision trees, it uses a Gini index to select n random features and finds the best optimal point. New observation is fed to the classification trees for predicting the target value of new data instances. Each classification tree is counted on the performance of the number of predictions for a class. The predicted class with the most votes is returned as a class label [25, 26, and 35].

3.5.5 Logistic Regression:

LR is a typical case of a classification technique that depends on the categorical class labels. The logistic regression classification model is a simplified classification model that computes predicted rates based on the likelihood of a certain outcome response variable. Here, the relationship between the dependent (labels - Y) and independent (features - X) variables is forecasted as the classification process advances [7] [27].

3.5.6 Feed-Forward Deep Neural Network:

This FF-DNN is the simplest type of artificial neural network; it has many applications in machine learning. The information flows only in forward direction, from the inputs to the output nodes, passing through any hidden nodes (if any). The numbers of input, hidden, and output layers are decided based on the experimentation. The input layer will have nodes equal to the number of features in the input data typically seen in the training stage. The data further flows down the network through hidden layer along with the weights, inputs, and a bias component. Finally, the output layer receives an output, which is essentially anchored through some of the activation functions [28, 29].

The General Algorithm for the proposed method is given below:

Algorithm – 1: Gender Classification of signature samples using different ML techniques

Input: Signature samples in a .CSV file.
Output: Gender classification

Step 1: Input features and labels from the signature database
Step 2: Perform feature engineering tasks on this dataset
Step 3:Perform the classification of each sample into Gender classes using the following ML techniques
i) Logistic regression, ii) KNN, iii) SVM, iv) Decision tree,
v) Random forest, vi) Feed-forward Deep neural network
Step 4: Compare the results of each technique using evaluation metrics
Step 5: Tabulate the results
Step 6: Stop

4. Experimental Results

i. Data splitting and model training

The core objective of this present research study is to classify gender based on signature features. For the classification task, six prominent ML algorithms have been used, namely KNN [21], SVM [23, 24], LR [27], DT [22], RF [26], and FF-DNN [29]. The efficiency of any ML algorithm is dependent on how the model is trained. Train_x and Test_x is used to train & test the variables, and also labels are trained as Train_y and tested as Test_y. The entire dataset is split into training and test sections in an 80:20 split ratio for the model training and testing. The model never used the test dataset until the evaluation of the model. Table 2 shows the samples used for training and test split. The research study also demonstrated the k-fold cross-validation [30] to avoid over-fitting [31] of the model during training.

ii. Hyperparameter Tuning

Hyperparameters cannot be trained from the data in almost all ML model architecture. However, when the Hyperparameters are not chosen appropriately, the whole model may fail to perform well [32]. It is because of this reason hyperparameter tuning is very crucial. For the KNN algorithm, the value of ‘k’ is essential and directly impacts the model’s accuracy. The sklearn library, by default, assigns the value of ‘k’ to 5. In the present study, a small python code was written to determine the accuracy of KNN model with value of the ‘k’ ranging from 1 to 20. The value which results in the highest accuracy was later used as the final value of ‘k’ before executing the model on the test data. Thus, ‘k’ = 8 resulted in better overall accuracy. For an SVM algorithm, adding kernels indicates a great way to improve the accuracy, so the model’s flexibility is further increased. The kernel was set into two categories, and the accuracy was calculated for each. The kernels chosen were RBF (with gamma=0.5, C=0.1) and poly (with degree=3, C=1) with decomposition strategy set to ‘ovo’ for both the kernels. As the results weren’t appreciating, the model was executed for cross_val scores defining Hyperparameters for each tiny subset of the overall dataset (defined in the form of folds). The DT classification model used both criterion parameters–Gini (accuracy=85%) and Entropy (accuracy=83%). The max_depth was set to 3. When the max_leaf_nodes=5 and max_features =13, both – Gini and Entropy variations resulted in the same accuracy (91%).

The RF has the following Hyperparameters measures as used in the algorithm: bootstrap=True, class_weight=none, criterion=‘gini’, max_depth=none, max_features=‘auto’, max_leaf_nodes=none, min_impurity_decrease=0.0, min_impurity_split=none, min_samples_leaf=1, oob_score=False, random_state=None, verbose=0, warm_start=False. For FF-DNN, the sequential model was used with one input layer (16, input_dim=63, activation= “relu”) and two hidden layers each with 32 and 64 neurons and activation= “relu” and one output layer (2, activation=“softmax”). The loss component was sparse_categorical_crossentropy as the softmax activation was used on the output layer. The Adam optimiser was used to deal with time complexity, and metrics were calculated using the ‘accuracy’ measure. The model was executed for 70 epochs, with the 66th epoch giving the highest accuracy of 98.2%.

Table 2. Training and Test split as 80:20 ratios for the given dataset before implementing an ML algorithm.

No of training samples: 80%	**Train_x**	**No of test samples: 20%**	**Test_x**	**Total**
	(428,66)		(107,66)	(535,66)
	Train_y		**Test_y**	
	(428,1)		(107,1)	(535,66)

Note: Train_x (80% of 535) indicates the 80% split from the original dataset for the training purpose, Test_x (20% of 535) is the 20% split reserved for the model validation. The corresponding labels stored at the variable ‘y’ are again split into 80% for Train_y and the remaining 20% for the Test_y to test the model on the unseen data.

5. Results and Discussions

The evaluation of results is dependent on various performance metrics as described in eq. 4-7. Those are i) True Positive (TP): the ML model correctly predicted the test sample into its true gender label (A female signature in the test sample is predicted with the label ‘female’). ii) True Negative (TN): the model correctly rejects attest sample from a definitive label (A male signature is not predicted as female signature). iii) False Positive (FP): A test sample is predicted with an incorrect label from the actual sample (Male signature is predicted with a label ‘Female’). iv) False Negative (FN): A predicted sample is incorrectly matched to its original label (A female signature is not recognized with a label ‘female’). v) A recall is to measure all true positive samples that model predicted correctly to the class; this indicates how many the model correctly predicted for the total samples of class 0. vi) A precision indicates the quality of the prediction, i.e., how many times the model correctly predicted a sample as class 0 out of all the total number of class 0 true samples. vii) F-Score is the average value between both recall and precision. iv) An accuracy is an actual number of samples that the model correctly classifies over the total number. viii) The macro average scores are calculated by considering the weighted mean for each R, P, and F for every predicted class without considering each label’s proportion. ix) The weighted average score is calculated by taking the product- sum of individual recall, precision, and f-score of each classified sample over the actual number of samples for the classification class. This is similar to the macro score except that the weighted score considers the proportion of individual labels. x) The micro average considers total TP, FP, and FN irrespective of the prediction made by the model for each class. Table 3 illustrates the performance achieved for each ML technique on the test dataset [33].

$$\boldsymbol{Recall(R)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{4}$$

$$\boldsymbol{Precision(P)} = \frac{\boldsymbol{TP}}{\boldsymbol{TP+FN}} \tag{5}$$

$$\boldsymbol{F-Score} = \boldsymbol{2} * \frac{\boldsymbol{P*R}}{\boldsymbol{P+R}} \tag{6}$$

$$\boldsymbol{Accuracy\ (A)} = \frac{\boldsymbol{TP+TN}}{\boldsymbol{TP+FP+TN+FN}} \tag{7}$$

The standard split method splits the dataset into training and test data. The accuracy is drawn from the original test samples. The user will not know the overall accuracy until the final execution. Suppose the accuracy is low, then the user must fine-tune the data and conduct the experiment. To rule out such a set-up, there is a provision to test the accuracy on the training data and finally run the model on the test data. Such an approach is called cross-validation, which once again splits the training data into 10-20% of validation data and run the model. Once the validation accuracy is high, the model is executed on the final test samples. This is repeated for various folds of the data and finally the accuracy average of each fold is calculated (the entire dataset is divided into equal number of partition called fold). To further authenticate the results obtained from the ML models, the authors have used a cross-validation metric to estimate the accuracy of the actual prediction obtained from each chosen algorithm. K-fold cross-validation (k=10) splits the entire dataset into ten folds. Additionally, 1-9 folds are randomly picked and used as a training set, and the remaining one-fold is to be preserved for the test set. Once the split is repeated ten times, cross_val_score is used to obtain the overall accuracy of individual folds. Further, cross_val_predict is used to get the average scores of all ten folds. See Table 4 for cross-validation scores of each fold. Figure 4 demonstrates the graphical representation of the overall accuracy for each model.

Table 3. The tabulation of various evaluation metrics on the test dataset for each ML method. The RF resulted with the highest accuracy, closely followed by KNN and DL methods.

	Class	**Precision**	**Recall**	**F_score**		**Macro**	**micro**	**Weighted**	**Accuracy**
KNN	0	0.91	0.71	0.80	P	0.83	0.81	0.99	0.81
	1	0.75	0.92	0.83	R	0.82	0.81	0.99	
					F	0.83	0.81	0.81	
SVM	0	1.00	0.33	0.80	P	0.80	0.66	0.90	Poly: 0.77 RBF: 0.49
	1	0.60	1.00	0.67	R	0.67	0.66	0.66	
					F	0.62	0.66	0.62	
LR	0	0.82	0.74	0.78	P	0.79	0.79	0.79	0.78
	1	0.76	0.83	0.79	R	0.79	0.79	0.79	
					F	0.78	0.79	0.78	
DT	0	0.88	0.94	0.91	P	0.91	0.91	0.91	0.91
	1	0.94	0.87	0.90	R	0.91	0.91	0.91	
					F	0.91	0.91	0.91	
RF	0	0.94	0.93	0.93	P	0.93	0.93	0.93	0.93
	1	0.93	0.94	0.93	R	0.93	0.93	0.93	
					F	0.93	0.93	0.93	
FF-DNN	0	1.00	1.00	1.00	P	0.96	0.98	0.98	**0.982**
	1	0.97	0.96	0.96	R	0.97	0.98	0.98	
					F	0.96	0.98	0.98	

Note: C: Class; P: Precision; R: Recall; F: F-score; M: Macro; m: micro; W: Weighted; A: Accuracy

Table 4. Tabulation of accuracy for each ML method for each fold in cross-validation approach. The k value is 10, where 0-9 folds randomly serve as the training set, and the remaining one fold acts as a test set.

Algorithms	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
KNN	0.81	0.90	0.83	0.90	0.79	0.81	0.93	0.81	0.90	0.78
LR	0.83	0.74	0.83	0.83	0.79	0.90	0.83	0.72	0.85	0.85
SVM	0.86	0.81	0.83	0.79	0.83	0.88	0.86	0.76	0.80	0.85
DT	0.88	0.95	0.95	0.86	0.74	0.88	0.88	0.86	0.92	0.85
RF	0.90	0.90	0.97	0.90	0.93	0.97	0.86	0.83	0.95	0.95
FF-DNN	0.97	0.98	0.98	0.97	0.98	0.98	0.98	0.98	0.98	**0.98**

As observed, the feed forward DNN performs better than other methods, followed by RF.

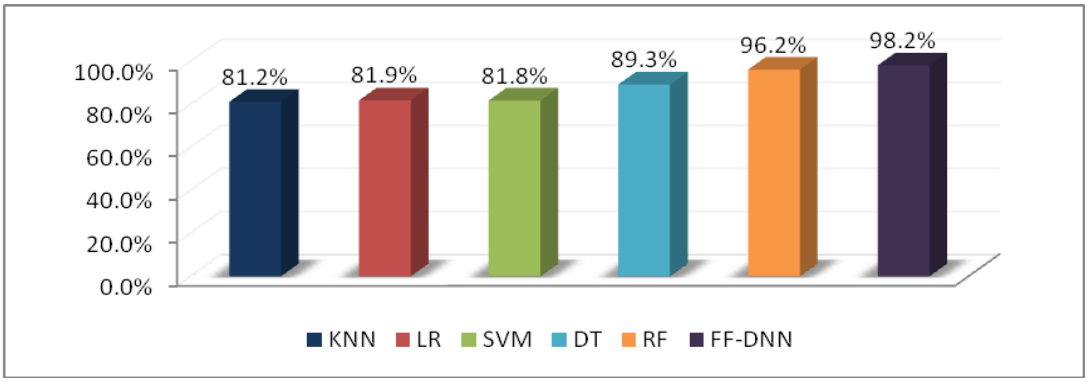


Fig. 2. Graphical representation of the overall 10-fold cross-validation accuracy obtained. The average accuracy for these scores is calculated and presented in the pictorial representation.

Comparative Analysis:

Table 5 shows a summary of the proposed work in relation to other methods discovered in related study. This comprises datasets, features/methodology, classification and results. It is observed that the proposed method is yielded a comparatively higher accuracy for gender detection than other methods (Table 5).

Limitations:

The research work has not concentrated on the pre-processing of signature images before the individual identification of the subject. A CNN model could have been used to imprint the feature extraction task on the signature obtained from the subjects. This approach could have helped build a .CSV file based on the sensor, match, feature and decision level parameters. Further, the latent space representation could have been used to suppress the features using an auto encoder [34]. In this case, the number of features could have been reduced (rather than 66 features). No additional strategies are used for the data augmentation task. The collected features were forecasted for the classification task. With data augmentation, the training of new algorithms that are based on ML models could have provided interesting patterns, which could alleviate the segment classification based on other categories such as identification of neurological disorders and others

Future enhancements:

The future work for this research study includes the following: (i) analyze these signature features to assess any human neurological disorder using ML algorithms. (ii) Study the novel multi-class signature-based problems, which could be in terms of age prediction and semantic language identification using signature features, (iii) extend the dataset size by collecting more handwritten signature samples to improve the results obtained with certain demographic information.

Table 5. Comparative Analysis

Authors	Features	Database	Classifier	Result
Yasemin Bay Ayzeren et al. [7]	Online signatures with time dependent features	804 handwriting and 8040 handwritten signature	Random Forest	58% for handwriting 50% for handwritten signature
Sura AbdAli et al.[8]	MRMR features	TCDatabase	Global Classifier	6.25% error rate for gender, 2.6% for Age
Thameur Dhieb et al. [9]	Fuzzy and Extended Beta-Elliptic Model	IBM_UB_1, ADAB	Deep Neural Network	96.90% for IBM_UB_1 98.25% for ADAB dataset
Marcus Liwicki et al. [10]	29 dynamic features	IAM-OnDB	GMM	67.06% for gender 84.66% for Handedness
Proposed Method	**66 dynamic signature features**	**535 (282 male and 253 female) online handwritten signatures**	**FFDNN**	**98.2%**

Table 1. System Generated Features

Features	Name	Description	Features	Name	Description
f1	N_STROKES	Number of down-strokes	*f32*	N_ASPECT_RATIO	Aspect ratio
f2	T_CONTACT	Time pen was down	*f33*	N_XTURNS	Number of X turns
f3	T_AIR	Time between strokes	*f34*	N_YTURNS	Number of Y turns
f4	T_TOTAL	Total time from first pen-down to last pen-up	*f35*	N_SPEEDCOR	Horizontal/vertical speed correlation
f5	N_PENSTOP_STA	No. of pen-stops found at the start of strokes	*f36*	V_RMS_SPEED	RMS pen speed
f6	N_PENSTOP_MID	No. of pen-stops found within strokes	*f37*	V_RMS_ACCEL	RMS pen acceleration
f7	N_PENSTOP_END	No. of pen-stops found at the end of strokes	*f38*	G_VEXTREMES	The angle between vertical extremes
f8	D_START_POS	Distance from start of signing line to first point	*f39*	G_SUM_ROTATION	Sum of rotations
f9	T_SUM_SEGSTA	Sum of times strokes started	*f40*	G_BASE_GRAD	The baseline gradient
f10	T_SUM_SEGEND	Sum of times strokes ended	*f41*	G_TOP_GRAD	The topline gradient
f11	D_WIDTH	Width of inked image	*f42*	N_EVENTS	Number of events
f12	D_HEIGHT	Height if inked image	*f43*	T_SUM_EVT_DUR	Sum of event durations
f13	D_WIDTH_OA	Overall width including pen-up positions	*f44*	D_SUM_EVT_POS	Sum of event positions
f14	D_HEIGHT_OA	Overall height including pen-up positions	*f45*	T_SUM_EVT_TIMES	Sum of event times
f15	D_UP_DIST	Pen-up distance	*f46*	T_SUM_MAX_EVT_DUR	Sum of maximum event durations
f16	D_DOWN_DIST	Pen-down distance	*f47*	T_SUM_MIN_EVT_DUR	Sum of minimum event durations
f17	D_X_POS	Positive X Down Distance	*f48*	T_SUM_MAX_EVT_TIMES	Sum of maximum event times
f18	D_Y_POS	Positive Y Down Distance	*f49*	T_SUM_MIN_EVT_TIMES	Sum of minimum event times
f19	D_X_NEG	Negative X Down Distance	*f50*	D_SUM_MAX_EVT_POS	Sum of maximum event positions
f20	D_Y_NEG	Negative Y Down Distance	*f51*	D_SUM_MIN_EVT_POS	Sum of minimum event positions
f21	D_NET	Net distance from start to end	*f52*	T_SUM_1ST_EVT_DUR	Sum of first event durations
f22	D_SUM_SEGSTA	Sum of stroke start distances(including up travel)	*f53*	T_SUM_LST_EVT_DUR	Sum of last event durations
f23	D_SUM_SEGEND	Sum of stroke end distances(including up travel)	*f54*	N_MCP_XD	x displacement
f24	D_CENTROID_X Coordinate	X coordinate of centroid	*f55*	N_MCP_YD	y displacement
f25	D_CENTROID_Y Coordinate	Y coordinate of centroid	*f56*	N_MCP_OD	overall displacement
f26	D_BASE_POS	The baseline position	*f57*	N_MCP_XS	x speed
f27	D_TOP_POS	The top line position	*f58*	N_MCP_YS	y speed
f28	D2_BOUNDING	Area of bounding rectangle	*f59*	N_MCP_OS	overall speed
f29	D2_NET	Net-area	*f60*	N_MCP_XA	x acceleration
f30	D2_RUBBER	Rubber-band area	*f61*	N_MCP_YA	y acceleration
f31	F_PRESSURE	Average pen force in device units	*f62*	N_MCP_OA	overall acceleration

6. Conclusion

In the present research study, the authors have focused on using six different ML algorithms for performing binary classification in gender identification using the numeric feature of the signature dataset. The proposed comparative study uses simple yet efficient models in terms of training speed and accuracy of the test dataset. Out of the chosen six algorithms, the FF-DNN performed better, with an accuracy of 98.2% on the overall classification task.

Author contributions

Shivanand S. Gornale[1]: Conceptualization, Methodology **Sathish Kumar*[2]:** Dataset Preparation, Writing-Original draft preparation. **Rashmi Siddalingappa[+2]:** Visualization, Investigation, Software, Validation. **Arjun Mane[3]:** Field study Writing-Reviewing and Editing.

Conflict of interest

The authors declare that the present study has no conflict of interest.

References

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

[2] Hiremath, P.S., & Hiremath, M. (2013). Radon Transform And Symbolic Linear Discriminant Analysis Based 3d Face Recognition Using KNN And SVM. DOI: http://warse.org/pdfs/2013/ijacst022122013.pdf

[3] Travieso, C. M., Alonso, J. B., Vasquez, J. L., Dutta, M. K., & Singh, A. (2017). Applying forensic features on writer identification. In 2017 4th International Conference on Signal Processing and Integrated Networks, SPIN 2017 (pp. 572–577). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/SPIN.2017.8050015

[4] Zois, E. N., Theodorakopoulos, I., Tsourounis, D., & Economou, G. (2017). Parsimonious Coding and Verification of Offline Handwritten Signatures. In IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (Vol. 2017-July, pp. 636–645). IEEE Computer Society. https://doi.org/10.1109/CVPRW.2017.92

[5] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., & Ortega-Garcia, J. (2021). DeepSign: Deep On-Line Signature Verification. IEEE Transactions on Biometrics, Behavior, and Identity Science, 3(2), 229–239. https://doi.org/10.1109/TBIOM.2021.3054533

[6] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., & Ortega-Garcia, J. (2019). Do you need more data? the DeepSignDB on-line handwritten signature biometric database. In Proceedings of the International Conference on Document Analysis and Recognition, ICDAR (pp. 1143–1148). IEEE Computer Society. https://doi.org/10.1109/ICDAR.2019.00185

[7] Sharma, V., Bains, M., Verma, R., Verma, N., & Kumar, R. (2021). Novel use of logistic regression and likelihood ratios for the estimation of gender of the writer from a database of handwriting features. Australian Journal of Forensic Sciences. https://doi.org/10.1080/00450618.2021.1956587

[8] Bay Ayzeren, Y., Erbilek, M., & Celebi, E. (2019). Emotional State Prediction from Online Handwriting and Signature Biometrics. IEEE Access, 7, 164759–164774. https://doi.org/10.1109/ACCESS.2019.2952313

[9] AbdAli, S., & Putz-Leszczynska, J. (2014). Age and gender-invariant features of handwritten signatures for verification systems. In Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2014 (Vol. 9290, p. 929021). SPIE. https://doi.org/10.1117/12.2075160

[10] Dhieb, T., Njah, S., Boubaker, H., Ouarda, W., Ayed, M. B., & Alimi, A. M. (2018). An extended beta-elliptic model and fuzzy elementary perceptual codes for online multilingual writer identification using deep neural network. ArXiv. Computer Vision and Pattern Recognition.

[11] Liwicki, M., Schlapbach, A., Loretan, P., & Bunke, H. (2007). Automatic detection of gender and handedness from on-line handwriting. Proc. 13th Conf. of the Graphonomics Society.

[12] Li, N., Liu, J., Li, Q., Luo, X., & Duan, J. (2016). Online signature verification based on biometric features. In Proceedings of the Annual Hawaii International Conference on System Sciences (Vol. 2016-March, pp. 5527–5534). IEEE Computer Society. https://doi.org/10.1109/HICSS.2016.683

[13] Durrani, M. Y., Khan, S., & Khalid, S. (2019). VerSig: a new approach for online signature verification. Cluster Computing, 22, 7229–7239. https://doi.org/10.1007/s10586-017-1129-4

[14] Kutzner, T., Pazmiño-Zapatier, C. F., Gebhard, M., Bönninger, I., Plath, W. D., & Travieso, C. M. (2019). Writer identification using handwritten cursive texts and single character words. Electronics (Switzerland), 8(4). https://doi.org/10.3390/electronics8040391

[15] Vorugunti, C. S., Mukherjee, P., & Pulabaigari, V. (2020). Online Signature Profiling using Generative Adversarial Networks. In 2020 International Conference on COMmunication Systems and NETworkS, COMSNETS 2020 (pp. 894–896). Institute of Electrical and Electronics EngineersInc.https://doi.org/10.1109/COMSNETS48256.2020.9027369

[16] Zhou, Y., Zheng, J., Hu, H., & Wang, Y. (2021). Handwritten Signature Verification Method Based on Improved Combined Features. Applied Sciences, 11(13), 5867. https://doi.org/10.3390/app11135867

[17] Zhang, X. Y., Xie, G. S., Liu, C. L., & Bengio, Y. (2017). End-to-End Online Writer Identification with Recurrent Neural Network. IEEE Transactions on Human-Machine Systems, 47(2),285-292. https://doi.org/10.1109/THMS.2016.2634921

[18] S, R., & M, H. (2017). Determining the Degree of Knowledge Processing in Semantics through Probabilistic Measures. International Journal of Information Technology and Computer Science, 9(7), 35–41. https://doi.org/10.5815/ijitcs.2017.07.04

[19] Gornale, S., Patil, A., & Hangarge, M. (2021). Palmprint Biometric Data Analysis for Gender Classification Using Binarized Statistical Image Feature Set (pp. 157–167). https://doi.org/10.1007/978-981-16-1681-5_10

[20] R, K., Patil, A., & Gornale, S. (2019). Fusion of Features and Synthesis Classifiers for Gender Classification using Fingerprints. International Journal of Computer Sciences and Engineering, 7(5), 526–533. https://doi.org/10.26438/ijcse/v7i5.526533

[21] Siddalingappa, R., & Kanagaraj, S. (2022). K-nearest-neighbor algorithm to predict the survival time and classification of various stages of oral cancer: a machine learning approach. F1000Research, 11, 70. https://doi.org/10.12688/f1000research.75469.1

[22] Safara, F., Mohammed, A. S., Yousif Potrus, M., Ali, S., Tho, Q. T., Souri, A., … Hosseinzadeh, M. (2020). An Author Gender Detection Method Using Whale Optimization Algorithm and Artificial Neural Network. IEEE Access, 8, 48428–48437. https://doi.org/10.1109/ACCESS.2020.2973509

[23] Gornale, S.S., Patil, A., & Prabha (2016). Statistical Features Based Gender Identification Using SVM. *International Journal for Scientific Research and Development*, 241-244.

[24] Rashmi, S., Hanumanthappa, M., & Jyothi, N. M. (2016). Text-to-Speech translation using Support Vector Machine, an approach to find a potential path for human-computer speech synthesizer. In Proceedings of the 2016 IEEE International Conference on Wireless Communications, Signal Processing and Networking,

WiSPNET 2016 (pp. 1311–1315). Presses Polytechniques Et Universitaires Romandes. https://doi.org/10.1109/WiSPNET.2016.7566349

[25] Singh, A., N., M., & Lakshmiganthan, R. (2017). Impact of Different Data Types on Classifier Performance of Random Forest, Naïve Bayes, and K-Nearest Neighbors Algorithms. International Journal of Advanced Computer Science and Applications, 8(12). https://doi.org/10.14569/ijacsa.2017.081201

[26] Hussein, S., Farouk, M., & Hemayed, E. S. (2019). Gender identification of egyptian dialect in twitter. Egyptian Informatics Journal, 20(2), 109–116. https://doi.org/10.1016/j.eij.2018.12.002

[27] Shiva Shankar, Raghaveni, J., Rudraraju, P., & Vineela Sravya, Y. (2020). Classification of gender by voice recognition using machine learning algorithms. Journal of Critical Reviews, 7(9), 1217–1229. https://doi.org/10.31838/jcr.07.09.222

[28] S, S., N, K., & Natarajan, S. (2011). Feed Forward Neural Network Based Eye Localization and Recognition Using Hough Transform. International Journal of Advanced Computer Science and Applications, 2(3). https://doi.org/10.14569/ijacsa.2011.020318

[29] Fayyaz, M., Yasmin, M., Sharif, M., & Raza, M. (2021). J-LDFR: joint low-level and deep neural network feature representations for pedestrian gender classification. Neural Computing and Applications, 33(1), 361–391. https://doi.org/10.1007/s00521-020-05015-1

[30] Wong, T. T., & Yeh, P. Y. (2020). Reliable Accuracy Estimates from k-Fold Cross Validation. IEEE Transactions on Knowledge and Data Engineering, 32(8), 1586–1594. https://doi.org/10.1109/TKDE.2019.2912815

[31] Berrar, D. (2018). Cross-validation. In Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics (Vol. 1–3, pp. 542–545). Elsevier. https://doi.org/10.1016/B978-0-12-809633-8.20349-X

[32] Probst, P., Boulesteix, A. L., & Bischl, B. (2019). Tunability: Importance of hyperparameters of machine learning algorithms. Journal of Machine Learning Research, 20.

[33] Rashmi, S., & Hanumanthappa, M. (2017). Qualitative and quantitative study of syntactic structure: a grammar checker using part of speech tags. International Journal of Information Technology (Singapore), 9(2), 159–166. https://doi.org/10.1007/s41870-017-0016-9

[34] Siddalingappa, R., & Kanagaraj, S. (2021). Anomaly Detection on Medical Images using Autoencoder and Convolutional Neural Network. International Journal of Advanced Computer Science and Applications, 12(7), 148–156. https://doi.org/10.14569/IJACSA.2021.0120717

[35] Gornale, S. S., Kumar, S., & Hiremath, P. S. (2021). Handwritten Signature Biometric Data Analysis for Personality Prediction System Using Machine Learning Techniques. Transactions on Machine Learning and Artificial Intelligence, 9(5), 1–22. https://doi.org/10.14738/tmlai.95.10808

[36] Wacom. [Online]. Available: http://www.wacom-europe.com

[37] Gornale, S., Kumar, S., Siddalingappa, R., & Hiremath, P. S. (2022). Survey on Handwritten Signature Biometric Data Analysis for Assessment of Neurological Disorder using Machine Learning Techniques. Transactions on Machine Learning and Artificial Intelligence, 10(2), 27–60. https://doi.org/10.14738/tmlai.102.12210