Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 2, No.1 : 2024
ISSN : **1906-9685**

GENDER HANDWRITING IMAGE IDENTIFICATION USING ENHANCED NAÏVE BAYES CLASSIFIER

Murugan A Associate Professor & Head, PG & Research, Department of Computer Science, Dr. Ambedkar Government Arts College (Autonomous) Affiliated to University of Madras, Chennai, India Email id: amurugan1972@gmail.com
Umadevi T P Assistant Professor, Department of Computer Science, JBAS College for Women (Autonomous), Chennai, India Email id: umashiva06@gmail.com

ABSTRACT
Generally, handwritings initialize the peculiarity of each person. Handwritings acts as a proof mainly for certain field such as court-of-law, archaeology, psychology, forensics, biometrics, historic document analysis and several other fields too. In handwriting the letter formation, neatness, uniformity, spacing and alignment of letters and sentences varies from person to person. A male possess one handwriting style, whereas a female posses a varied handwriting style. When it comes to transgender, there are similar variations Hence, for handwriting-based gender classification issue several machine learning classifiers are applied that are not efficient. This paper proposes an enhanced Naïve Bayes classifier for handwriting gender classification. A training dataset implementation is applied collecting 1000 participants' handwritten samples in English. This system gives an accuracy of 85%.
Keywords- Naive Bayes Theorem, Gender Classifier, Deep learning approach

INTRODUCTION
In recent times psychological sciences have proved the importance of handwriting in each ones personal lives [1]. Handwriting helps in personal development, retaining information, framing topics and knowing the characters of people. Generally, based on the handwriting the gender of a person can be predicted. Gender classification by handwriting helps in several ways in detecting the criminals, hackers, selecting profession, signature fraudulent prediction etc.., though several machine learning approaches are proposed still gender classification by handwriting is being a challenging task in extracting accurate results. The basic structure of analyzing handwriting indulges handwriting recognition, identifying writer and signature verification. Arabic is one of the most important worldwide languages employed in documenting sources. The language 'Arabic' poses several challenges because of totally different handwriting designs and changeable letter shapes that appreciate different spatial localities [2]. Around twenty-seven languages use the Arabic alphabet. Arabic is the language of quite 420 million people around the world, creating it the sixth most used spoken language. Hence, Arabic language handwriting recognition is considered to be more challenging for computers vision to track the letters than other languages [3].
Sokic[4] . used ensemble classifiers for improvising the gender classification by handwriting. Also, a wavelet-form gender detection automation approach used by akbari in [5] predicting the off-line handwritten documents were used. LeCun and coworkers [6]stated 'a multilayer network' termed as LeNet-5 derived from the CNN architecture categorizing handwritten numbers [7]. The recent approaches [8],[9] produced a high-level combination theories that improves the handwriting recognition systems performance.

The paper proposed an enhanced feature extraction and improved classification method to differentiate the genders using handwritings. Further as an added augment a deep learning approach to re-solve the human errors are applied.

The proposed theory guarantees simple, efficient, and accurate results. Also, the proposed system present a fairly huge and diverse training dataset accounted from kaggle containing participants more than 1000. The results are comparable to the reports of previous methods, and the proposed method provides better accuracy rates from the dataset.

LITERATURE SURVEY
Several approaches were tried to enhance the gender classification by handwriting. Some of the novel approach used for handwriting recognition (numbers, characters and words) is stated in many articles which are to be discussed.
Angel Morera et al. [10], Uses CNN approach, that specifies three categorizations namely handedness, gender and joint gender-and-handedness for solving the demographic problems based upon the deep neural networks (DNN).
S. K. Jemni et al. [11] produced a hybrid combination of cooperative classifiers together for automatic handwriting recognition. The two main categories of combination Techniques are feature fusion and decision fusion. Feature fusion provides a combination of features to single feature vector by using single classifier [13]. The second is decision fusion provides integration on varied classifiers for making decision. X. Liu, [12] uses multiple sliding windows related feature extraction for proposing a superior recognition rate.
S. A. Azeem and H. Ahmed in [14], Proposed a pyramid histogram of gradients for HoG feature extraction achieving localization along with markov process. Sokic et al. [16] initiates a set of differentiators tangent and curvature intimations and 4 differentiators to differentiate between male and female handwritings.

PROPOSED WORK
3.1 APPROACH CATEGORIZATION
The proposed approach is based on segmenting the writing of the people as words to letters. Initially the system regulates a feature extraction process, a probability classifier, naïve bayes theorem implementation, pattern matching, fuzzy calculation and gender classification by handwriting produced in figure 1.

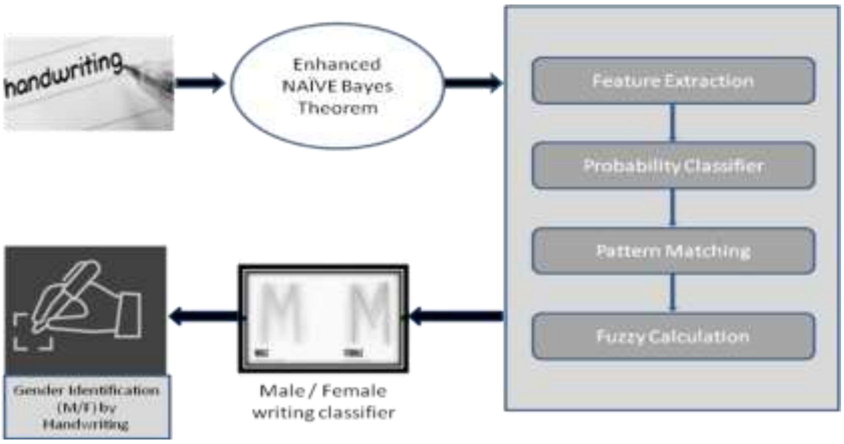


Figure 1: System architecture on Gender Identification by Handwritten Image

Feature extraction targets to locate a representation of handwriting images that initializes differentiation among male and female writers. The proposed system employs the segmentation information to differentiate between the two gender classes. The feature extraction process has been recruited for solving several recognition problems. The goal of the paper is not only to propose original features but also to provide enhancement on the classification rates using the classifiers. A probability check is made where the chances of male and chances of female writings are predicted and separated for further validation process.

In this paper, a new enhanced Naïve Bayes classifier used for the model distribution. Here the handwriting is processed for verification and identification process using the writing style, characters and formations the classifications are made. This algorithm classifies the gender in an accurate manner. In pattern matching the previous records are used and the patterns are matched with the present deliberated handwriting. If patter matches then clustering is done. A Fuzzy interference system is set for calculation/predicting the handwriting. A fuzzy value is set for each character. The variations and formations of each word are analyzed and at last an accurate prediction is made. The gender classification model helps to classify the prediction of male handwriting, female handwriting and transgender handwriting separately. After, performing the gender classification the final results are provided. Simultaneously by implementing a deep learning approach the human errors and fraud activities are also identified and rectified.

Handwriting Image is depicted in figure 2.

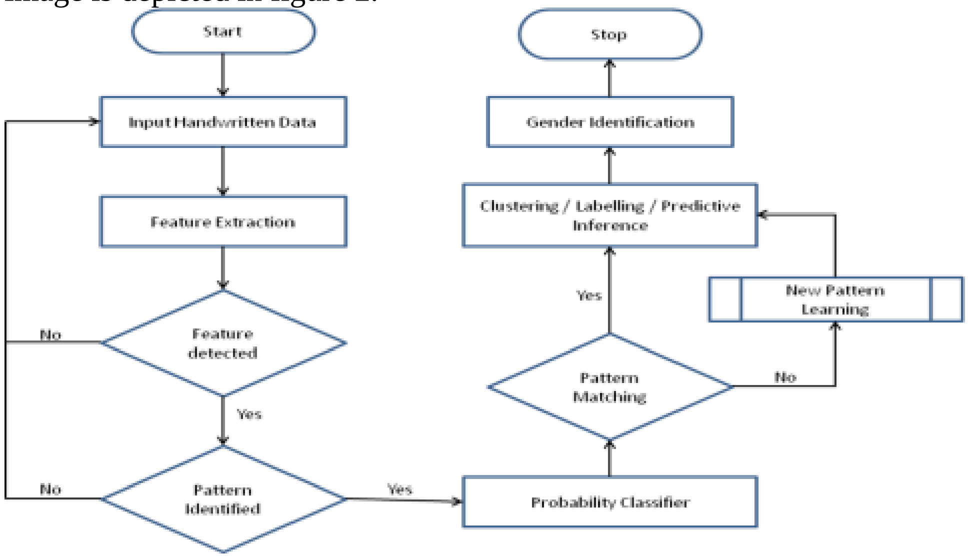


Figure 2: Process flow of Gender Identification by Handwriting Image

The process flow of Gender Identification that includes male/female/transgender by Handwriting Image is depicted in figure 2.The features are extracted from the image and checked for the pattern match. If the pattern does not match it will check for new pattern learning.

3.2 An hybrid Algorithm for Gender Handwriting Image Identification

In the below algorithm a handwriting dataset is provided as input and described as $HW_{ds.}$ The training dataset is initialized fro handwriting pre-processing and extracting the feature de and fe. When the features are detected then the process is regulated for pattern matching Pid. By using the probability check $P_r = x \in X$ a classification of gender male/female writing is made. If patterns are identified pm, then naïve bayes theorem is applied. Then, the process of clustering c, labelling, predictive interference, fuzzy calculation and finally gender identification is performed Gi. If the patterns are not matched new pattern learning p1 is executed and the process is done again for performing gender identification and classification by handwriting.

Gender Handwriting Image Identification Using Enhanced Naïve Bayes Classifier
Algorithm: Enhanced Naive Bayes Algorithm
Input: handwritten document or handwritten documents
Output: Identifying Male or Female
Begin
Read the hand written documents
doc=0

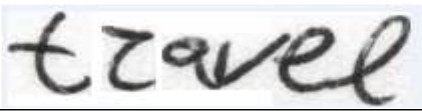
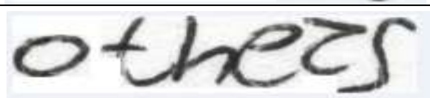
While (doc = true)
If Feature detected then
Perform Pattern identification with Training dataset
If Pattern identified then
Perform Probabilistic Classifier Perform Pattern Matching with training dataset
If Pattern Matched
Perform Clustering, Labelling, Predictive inference & fuzzy
If Gender Identified is Male
Increment Male by 1
Else
Increment Fe-Male by 1
End if
Else
Learn New Pattern learning using Deep learning approach.
Continue Loop
End if
End if
Else
Proceed to next loop
End If
If Male > Fe-Male then
Print Identified Gender is "Male"
Else
Print Identified Gender is "Fe-Male"
End if

End Loop
EndWhile
End

RESULT ANALYSIS

A training data set containing 1000 participants handwriting images are considered. Through the feature extraction the male formations and female based formations are collected. Then, sent to the pattern matching in case of matches found the collected samples are clustered and assigned as female image and male image. Using the Naïve Bayes classifier the gender classification is performed classified as male subject and female subject

Table 1 Handwriting Images Samples

Male Subject	**Female Subject**
	
	
	
	

In the table1 and Figure 3 the collected male subjects and female subjects that are stated according to the letter formation are shown. In the table 2 the parameters mean, accuracy rate, energy, contrast, correlation rate and homogeneity are produced along with rate for both male and female subjects.

Table 2 - Sample Feature extraction values for male and female handwriting images

Parameters	**Image 1 (Male Subject)**	**Image 2 (Female Subject)**
Mean value	209.38	189.11
Accuracy	**75%**	**85%**
Performance	**85%**	**100%**
Contrast	0.98179	0.93028
Correlation Rate	0.87841	0.90394
Energy	0.41111	0.26612
Homogeneity	0.8517	0.8107

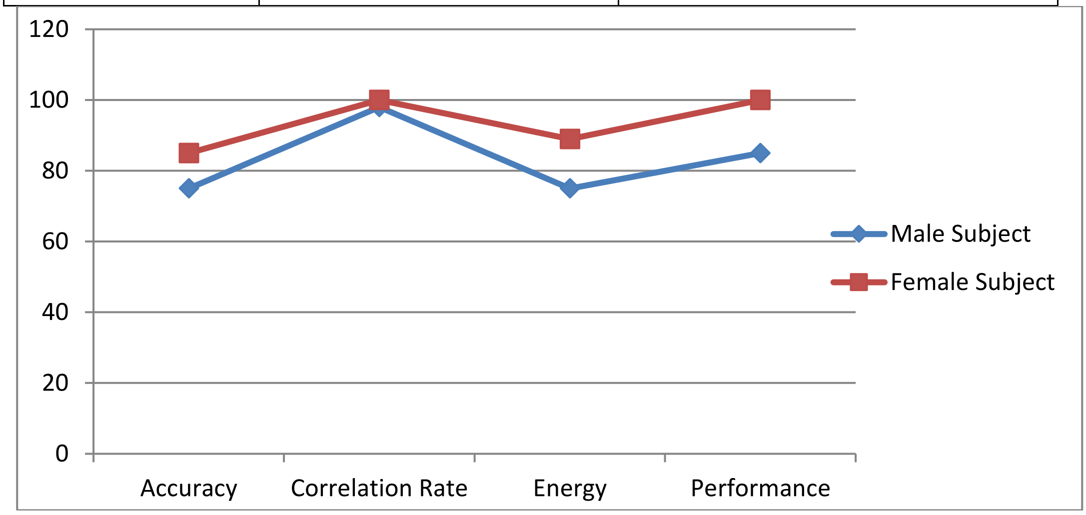


Figure 3: Performance of the Enhanced Naive Bayes Classifier Method

The samples stored in the data base are taken for comparative for analysis. Performance is evaluated in (Figure 4) by comparing the samples with the proposed Enhanced naive Bayes theorem and Existing SVM, CNN and ensemble theorem. According to the comparison performing analysis the enhance Naïve Bayes algorithms provides more accurate results compared to the other algorithms. Training set is compared with data already stored.

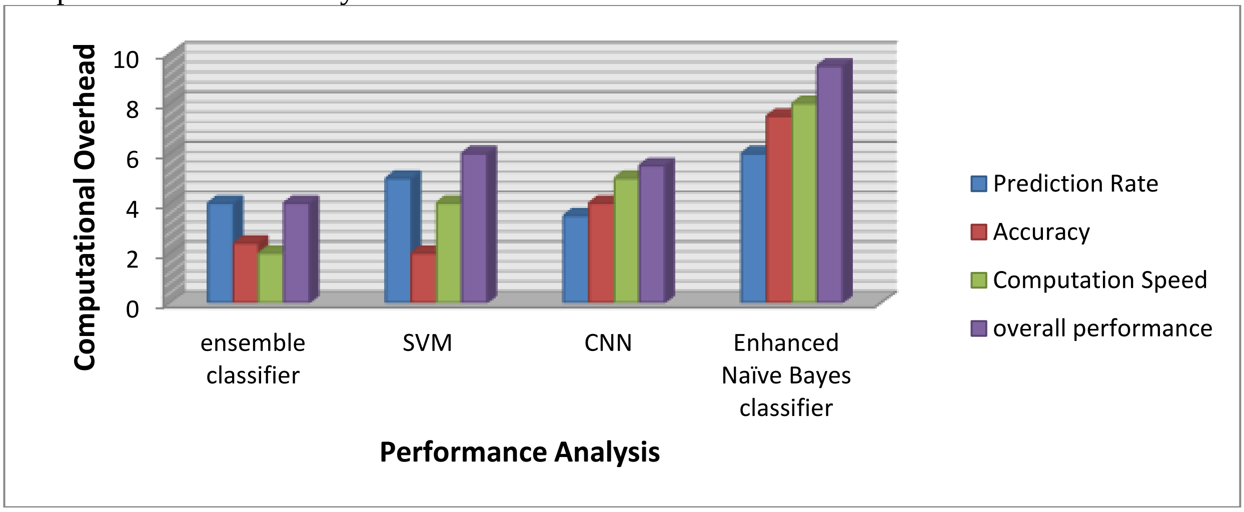


Figure 4: Performance Analysis Comparing Proposed and existing systems

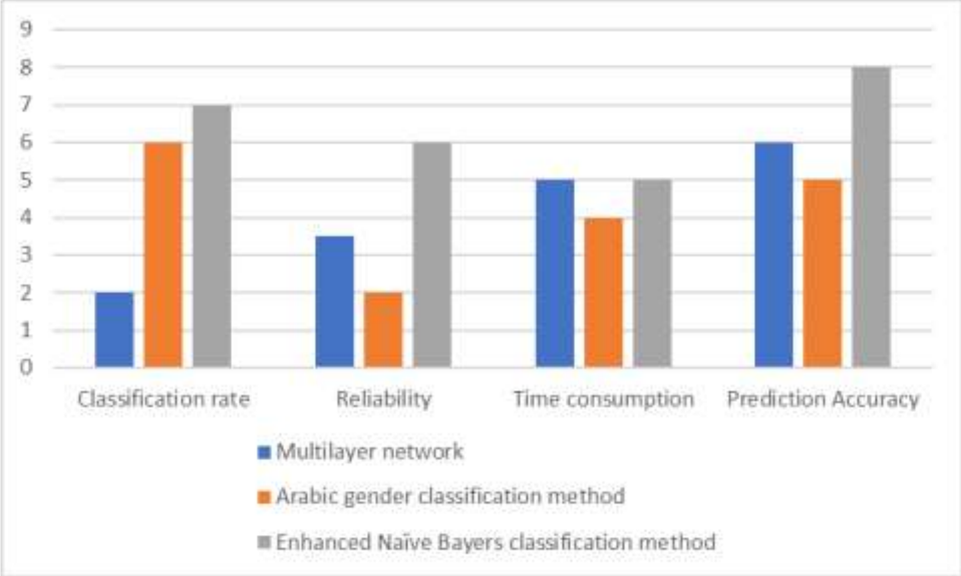


Figure 5: Performance Evaluation comparing various migration models

In Fig. 5, the comparison of Enhanced Naïve Bayes Classifier is made with the existing system Multilayer NW and Arabic Gender classification method and the Proposed system is proved to be best in Gender classification providing speed, classification, reliability, minimal time and Accurate prediction.

CONCLUSION

The system proposes an effective enhanced naïve bayes theorem for classifying the handwriting performing the identification too. The technique mainly relies on extracting a set of features and training classifiers using writing samples of male and female writers. Different classifiers are then combined performing feature extraction, pattern matching, probability, fuzzy interference and gender identification. Using the training dataset the accurate results for handwriting recognition and gender classification model is projected.

The future work focus on predicting extra-attributes of the handwriting images such as age, languages, handedness etc.. Hence, extra languages are tried with big dataset containing more samples.

REFERENCES

[1] P. A. Mueller and D. M. Oppenheimer, “The pen is mightier than the keyboard: Advantages of longhand over laptop note taking," *Psychol. Sci.*, vol. 25, no. 6, pp. 1159_1168, 2014.

[2] A. Hasasneh, N. Salman, and D. Eleyan, “Towards of_ine Arabic handwritten character recognition based on unsupervised machine learning methods: A perspective study," *Int. J. Comput. Academic Res. (IJCAR)*, vol. 22, pp. 329_349, Aug. 2019.

[3] I. A. Doush, F. Alkhateeb, and A. H. Gharaibeh, “A novel Arabic OCR post-processing using rule-based and word context techniques," *Int. J. Document Anal. Recognit.*, vol. 21, nos. 1_2, pp. 77_89, Jun. 2018.

[4] Sokic, E., Salihbegovic, A., Ahic-Djokic, M.,. “Analysis of off-line handwritten text samples of different gender using shape descriptors”, *Proc. of IX International Symposium on Telecommunications (BIHTEL)*, pp. 1–6. 2012

[5] Akbari, Y., Nouri, K., Sadri, J., Djeddi, C., Siddiqi, I, “Wavelet-based gender detection on off-line handwritten documents using probabilistic finite state automata”,*International Symposium on Networking*, pp 17–30 (2017).

[6] Y. LeCun et al., “Handwritten digit recognition with a back propagation network”, *Advances in neural information processing systems , Morgan Kaufmann Publishers*, pp. 396–404, 1990.

[7] S. Albelwi and A. Mahmood, “A framework for designing the architectures of deep convolutional neural networks,” *Entropy*, vol. 19, pp. 1–20, 2017.

[8] M. Awni, M. I. Khalil, and H. M. Abbas, “Deep-learning ensemble for of_ine Arabic handwritten words recognition," in *Proc. 14th Int. Comput. Eng. Syst. (ICCES)*, pp. 40_45,Dec. 2019,
[9] E. M. Hicham, H. Akram, and S. Khalid, “Using features of local densities, statistics and HMM toolkit (HTK) for of_ine Arabic handwriting text recognition," *J. Electr. Syst. Inf. Technol.*, vol. 4, no. 3, pp. 387_396, Dec. 2017
[10] S. K. Jemni, Y. Kessentini, S. Kanoun, and J.-M. Ogier, “Offline Arabic handwriting recognition using BLSTMs combination," in *Proc. 13th IAPR Int. Workshop Document Anal. Syst. (DAS)*, pp. 31_36, Apr. 2018.
[11] Kim, J., and Lee, S, "Gender Classification in Handwritten Documents UsingConvolutional Neural Networks." Pattern Recognition Letters, vol. 168, pp. 92-100, Mar. 2023.
[12] Nguyen, D. T., and Tran, M. D, "Lightweight Cryptographic Algorithms for Secure Cloud Storage: A Review." Journal of Cloud Computing: Advances, Systems and Applications, vol. 14(2), pp. 231-245. Apr. 2024.
[13] X. Liu, G. Meng, and C. Pan, “Scene text detection and recognition with advances in deep learning: A survey," *Int. J. Document Anal. Recognit., vol. 22, no. 2*, pp. 143_162, Jun. 2019.
[14] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, “A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11_26, Apr. 2017.