IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 25, Issue 2, Ser. II (Mar. – Apr. 2023), PP 26-31
www.iosrjournals.org

Nationality Identification using Handwriting

Sudeshna S
Computer Science & EngineeringVNRVJIET
Hyderabad, India sudeshna_s@vnrvjiet.in

Gopikrishna Gogineni
Computer Science & EngineeringVNRVJIET
Hyderabad, India gopichowdaryg23@gmail.com

Ramsai Rapole
Computer Science & EngineeringVNRVJIET
Hyderabad, India rapoleramsai4@gmail.com

Sai Kishore Reddy Regatte
Computer Science & EngineeringVNRVJIET
Hyderabad, India regatte.saikishore@gmail.com

Saiteja Rudroju
Computer Science & EngineeringVNRVJIET
Hyderabad, India saitejabablu94@gmail.com

Abstract—*Handwriting and signature biometrics, particularly in terms of identity recognition and verification have a long history. The process of handwriting analysis involvesexamining someone's writing. Graphology is the term used in science to describe handwriting analysis. It is a technique for extrapolating a person's personality and behaviour from his writing quirks. When crimes include people of different nationalities, it can be difficult for forensic investigation teams to pinpoint the crime. There are many uses for categorising handwriting according to factors like age, gender, and nationality. Investigations in forensics can be narrowed downto a particular type of writer with the aid of handwriting classification. This project proposes a new method for ethnicity(nationality) identification.*

Keywords: *Handwriting analysis, COLD features, Ethnicity identification, Nationality identification.*

Date of Submission: 17-03-2023 Date of Acceptance: 01-04-2023

I. INTRODUCTION

Gender, nationality, age, height, gait, and other characteristics are frequently used in biometric applications like iris and face recognition . This is due to the fact that trait prediction makes biometric methods more effective by making the problem less complex.Trait prediction also plays a significant part in forensic applications and security by aiding in the detection of questionable behaviours. However, it has been found that biometric-based systems lose accuracy when the input photographs are exposed to a public space. This is because biometric-based techniques have inherent flaws, such as sensitivity to outside influences, occlusion, deterioration, and uneven lighting effects.. Due to the intricate image processing necessary, it is also claimed that the methods are computationally expensive.As a result, handwriting analysis has caught the interest of academics, who are increasingly exploring outside of the box to predict features such as age, gender, country, and other traits in addition to more conventional traits like emotions. This is carried out to support forensic software and investigation teams. However, due to the enormous range of handwriting, ink, pen, paper, script, age, gender, and individual variances, reliable trait identification based on handwriting analysis is challenging. Because it makes it easier to identify crimes in which persons of different nationalities are engaged, this initiative is focused on nationality and ethnicity identification.

II. RELATED WORKS

The goal of Jungpil Shin et al., [1]'s machine learning project was to create a model that can classify, based on their handwriting, adults, and children. They used a sequential forward floating selection method for selection of parameters and the most efficient features. They combined the SFFS feature selection algorithm with the RF and SVM classification algorithms to separate adults and children into different categories. After selecting the common features from the sequential forward floating selection-SVM and SFFS-Random Forest classifiers, they applied the SVM and Random Forest classifiers for adults and children classification. Their findings suggested that Random Forest was a better classifier.

A thorough explanation of how to determine gender and age from handwritten papers is given in the Ching Y. Suen et al.,
[2] study. Age prediction systems are essential in the forensic and medical industries, and gender detection has applications in a wide range of fields, including psychology, document analysis, forensics, and many more.Transfer learning technique uses two advance CNN architectures Resnet and Googlenet to automatically extract features.It is observed that Resnet has higher accuracy in predicting the gender and age range classification due to skip connections to jump over layers.

Demographic handwriting classification has many applications in a variety of fields, including forensics, psychology, archaeology, and biometrics, according to Mina Rahmanian et al., [3] in their discussion of gender and handedness classification based on handwriting using CNN. They used KHATT (Arabic texts) and IAM (English texts)databases, along with three advanced CNNs, Xception, DenseNet201, and InceptionV3 to examine the effectiveness of deep CNNs in automatically classifying gender and handedness for two handwriting-based demographic problems. Additionally, it was investigated whether deep

CNNs could automatically categorise gender and handedness, two demographic issues based on handwriting.

A Parkinson's disease prediction method based on SVM and PCA was provided by Zhifei Xu et al., [4]. This method uses classifier modelling for the Meander repository and the spiral data set to classify & for identification of Parkinson's disease. Additionally, it makes full use of the sensor's six channels of data. They divided the data set by five using 5-fold cross-validation for the evaluation of the classification model's prediction performance with added precision and objectively and for achievement of a good rate of accuracy. This method has the potential to be used to make a clinical diagnosis of disease Parkinson’s.

A unique approach to writer identification and gender prediction using the Gurumukhi (Punjabi) script was put out by Shaveta Dargan et al., [5] using a combination of feature extraction methods and hybridized classification techniques.In this work, two feature extraction methods are taken into consideration. Both author identification and gender classification use SVM, MLP, KNN, Random Forest, and hybridizations of given classifiers. It has been claimed that classifier hybridization produced more accurate findings. By computing indicators like true and false positive rates (TPR and FPR), the authors revealed performance evaluation.

The handwritten signature is the biometric attribute that is most accepted for validating documents such as letters, contracts, wills, MOUs, etc. in daily life. Shivanand S. Gornale et al., [6] proposed project is based on the combination of statistical and textural information taken from the photographs of signatures. KNN, SVM and Decision Tree were used to predict gender. SVM is said to generate outcomes that are more accurate. The suggested approach is anticipated to be useful in the development of effective tools of computer vision for the forensic examination and authentication of documents consisting of handwritten signatures.

By analysing handwriting data gathered by pen-tablets, Shammi Akhtar et al., [7] suggested two systems for person identification and handedness prediction. Here, six features namely pressure, time, horizontal angle, x, y and vertical angle have been employed. These features are read from the excel file that consists of data stored by the pen and pad. To serve two separate aims, two systems have been developed. Random Forest and SVM classifiers are used to identify people, and SVM is used to identify a person's handedness. With more people using pen-tablets, this effort will make handwriting analysis using pen-tablet data more practical.

Handwriting based Personality Identification using Textural Features, Abdellatif Gahmousse et al., [8] proposed oBIFC technique for feature extraction.Five personality factors were used to assess each person, and the evaluation was treated as a multi-label classification issue.. On an average, oBIFC with RankSVM shows higher accuracy levels than previously existing methods with 58.6%.The classification performance can be affected by the imbalance of data distribution across the two sets used in the proposed system. The suggested system is also subjected to a 10-fold cross validation technique, which provides a more precise indicator of its efficacy.

In their discussion of handwritten individual characters analysis for human behaviour identification using graphology, Subhankar Ghosh et al., [9] proposed a system for identifying human behaviours using graphology.The system extracts different types of structural features from individual characters and categorise based on different parameters.The performance of features with rules defined shows better results than pre existing methods with SVM and CNN.Because the suggested system uses individual characters as the input for person behaviour recognition in this work, its performance suffers when someone writes touching characters.

Yasemin Bay Ayzeren et al.,[10] first presents a thorough analysis of a variety of unimodal and multimodal biometric datasets, with a focus on the availability, labels, content, and sample counts of handwriting biometrics as well as signature biometrics in this paper. With the newly proposed database, they have specially conducted a few experiments to demonstrate that it is viable to accurately forecast the state of emotion from both handwriting biometrics as well as signature biometrics. As a result, conclusions drawn from the emotion prediction are more often represented as broad suggestions of possibility than as firm statements. This paper raises several opportunities for further examination like prediction of emotions based on offline handwriting.

Handwriting identification using deep CNN method by Oka Sudana et al., [11] In this paper, using pre-trained CNN, they applied transfer learning to identify writers based on their handwriting. Using the IAM handwriting dataset, which includes 100 classes of writers, we trained the model. Pre-trained model VGG19 performs best when employing grayscale images rather than binary or invertedbinary images, according to the results of training and testing. Their research shows that handwriting images in sentence form can be used directly without the requirement for further feature extraction or segmentation techniques. The time it took to complete the training process is this study's drawback. To complete 100 epochs of the training process, it took close to 10 hours.

Abdeljalil Gatar et al., [12] described an effective approach to gender characterization based on handwriting using COLD and Hinge as features. This study utilized highly identifiable representations captured by curvature and contour information embedded in handwritten documents. Various combinations of selected features were examined with the SVM classifier. Through this study, they showed how both her SVM decision maxima are determined. B. For COLD and hinge features, handwriting-based gender detection can achieve high accuracy.

The discussion of personality identification from handwriting in Hastuti Fatimah et al., [13]'s study covered several personality reviews. They employed graphology analysis, which combines structural and symbol analysis techniques. Using the Convolutional NN classification, four specific characters (‘t’, ‘s’, ‘g’, and 'a') were examined. the features of margin, space among lines, between words, slope were subjected to multi-structure analysis. The accuracy of using CNN to recognise specific letter symbols was 98.03%, and the accuracy of using structural analysis ranged from 82.5 to 100%.

Adeljalil Gattal et al., [14] discussed an effective technique that makes use of oBIFs and uses the histograms and columns of oBIFs as characteristics is described for identifying gender from handwriting. SVM classifier is used to investigate various oBIF configurations. The proposed method performs better than the current approaches at both script dependent and script independent systems. In their follow-up research on this issue, they planned to look at how well oBIFs characterise other characteristics of people, such as age, nationality and disease recognition.

Somaya Al Maadeed et al., [15] discussed the classification of handwriting into age ,gender, and nationality using handwriting has wide range applications in forensics.Initially,images are binarized and classification is performed using Random forest classifier with Kernel discriminant analysis using spectral regression.The system uses several geometric features for classification in which only one feature outperforms in classification of age,gender, and nationality.Based on the results,random forest is mostly preferred for prediction of age and nationality whereas KDAis preferred for prediction of gender.

III. INVESTIGATION & FINDINGS OF PREVIOUS PAPERS

Table 1: Table showing different methodologies, pros, cons and the results obtained in this literature survey.

S.No	Title	Methodology	Pros/Cons	Year
1	Important Features Selection and Classification of Adult and Child from Handwriting Using Machine Learning Methods.	Random Forest SVM	Classification is done on both handwritten text and handwritten pattern databases. Random Forest takes care of null values and is too slow when the tree count is large.	2022
2	Transfer Learning toDetect Age From Handwriting	Using CNN architectures as Resnet andgooglenet and SVM	ResNet outperformed GoogleNet in accuracy, achieving a score of 69.7% as opposed to 61.1%. Resnet and Googlenet have manyhidden layers which take more computational time.	2022
3	Handwriting-based gender and handedness classification using convolutional neural networks	Convolutional Neural Networks. InceptionV3, and DenseNet201	Advanced Architectures were usedfor classification. Both of the algorithms are time-consuming and require high computational power.	2021
4	Handwritten dynamics classification of Parkinson's disease through support vectormachine and principal component analysis	PCA feature selection.SVM classifier.	PCA improves algorithm performance, reduces overfitting. SVM is memory efficient.77 % average accuracy rate using SVM.	2021
5.	Gender Classificationand Writer Identification System based on Handwritingin Gurmukhi Script.	MLP and SVM Curve fitting based feature extraction , Open Endpoint and Intersection based technique	Maximum accuracy rate of 87% is obtained. MLP computations are difficult andtime consuming	2021
6.	Handwriting Based Gender ClassificationUsing COLD and Hinge Features	COLD and Hinge feature.SVM Classifier	Best among both the feature extraction methods is considered. Unsuccessful classification resultswere observed in cases where hinge features were used.	2020
7.	Behavioral Biometric Data Analysis for Gender Classification Using Feature Fusion and Machine Learning	Decision tree and SVM Combination of statistical and texturalfeatures from the signature images.	Proposed technique can be helpful in the process of designing well runcomputer vision tools for forensic investigation and authentication of documents consisting of signatures. Decision tree is unstable.	2021
8	Handwriting based Personality Identification usingTextural Features	RankSVM,CVT(k-fold Cross Validation Technique)	oBIFC with RankSVM and CVTshows more accurate results thanwith existing systems SVM and CNN. Performance on one attribute out of five doesn't outperform in proposed methods	2020
9	Graphology based handwritten character analysis for human behaviour identification	Predefined rule based,SVM,CNN	Proposed features and rulesoutperform the other two approaches (SVM,CNN). Anticipates that letters in both lower- and uppercase will behave similarly.When	2020

			touching charactersare utilised, the proposed system's performance suffers.	
10	Prediction of emotionsform Handwriting	K fold cross validation,where k=4.SVM	Prediction is done on both onlineand offline datasets. Doesn't work for normal devices.	2020

11	Identification of handwriting scriptsusing deep CNN.	CNN VGG (Visual Geometry Group) pre-trainedmodel.	Taking binary image, grayscale image, and inverted binary imageas training dataset improves the accuracy of the model. Training process took almost 10hours to complete 100 epochs.	2020
12.	Analysis on Handwriting Using Pen-Tablet for Identification of Personand Handedness	SVM Random Forest	With more people using pen-tablets, this effort will make handwriting analysis using pen-tablet data more practical. Two different systems are created for person and person's handedness identification. Random forest requires more computational powerand resources.	2021
13	Personality Features Identification from Handwriting Using Convolutional Neural Networks	Graphology.CNN.	More accuracy of new data withvery few minutes in the training process. CNN automatically detects important features without humansupervision.	2019
14	Gender classificationfrom handwritten offline images	SVM classifier	The evaluated method performsbetter than the current trends inscripts that are dependent and independent. Effective only for gender classification and does not work efficiently for age,handedness andother classifications.	2018
15	Automatic prediction ofage, gender, and nationality in offline handwriting	Random Forest Classifier and KDA	The classification results for gender and age range classification outperforms previously existing methods with 73%(for gender) and 55%(for age). Random Forest takes care of null values and is too slow when the tree count is large.	2014

IV. CHALLENGES FACED BY EXISTING SYSTEM

In the existing systems of handwriting based classification there are very few systems that predict nationality. Various classifiers are used in existing systems such as Convolutional Neural Network architectures, Random Forest, Decision Trees and Support Vector Machine. CNN architectures, Inception v3 and densenet201 are time-consuming and require high computational power. Similarly, MLP computations are difficult and time consuming too. Random Forest takes care of null values but is too slow when tree count is higher. Decision trees are unstable even if one step of classification goes awry.

V. CONCLUSION

We provide an in-depth survey of the literature on handwriting-based predictions of several features, including gender, age range, handedness, and their corresponding accuracy on several datasets. In order for the models to be capable of handling problems and problem kinds that they have never seen before, more study is required, as is shown in this work, which includes all the models for handwriting-based prediction of various qualities. In this study, we also examined many datasets associated with this problem, their complexity, and the model accuracy on each individual dataset.

REFERENCES

[1]. Shin, J., Maniruzzaman, M., Uchida, Y., Hasan, M.A.M., Megumi, A., Suzuki, A. and Yasumura, A., 2022. Important Features Selection and Classification of Adult and Child from Handwriting Using Machine Learning Methods. Applied Sciences, 12(10), p.5256.

[2]. Najla, A.Q. and Suen, C.Y., 2022. Transfer Learning to Detect Age From Handwriting.

[3]. Rahmanian, M. and Shayegan, M.A., 2021. Handwriting-based gender and handedness classification using convolutional neural networks. Multimedia Tools and Applications, 80(28), pp.35341-35364.

[4]. Xu, Z. and Zhu, Z., 2021, April. Handwritten dynamics classification of Parkinson's disease through support vector machine and principal component analysis. In Journal of Physics: Conference Series (Vol. 1848, No. 1, p. 012098). IOP Publishing

[5]. Dargan, S. and Kumar, M., 2021, February. Gender Classification and Writer Identification System based on Handwriting in Gurumukhi Script. In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) (pp. 388-393). IEEE.

[6]. Gattal, A., Djeddi, C., Bensefia, A. and Ennaji, A., 2020, June. Handwriting based gender classification using cold and hinge features. In the International Conference on Imageand Signal Processing (pp. 233-242). Springer, Cham.

[7]. Gornale, S.S., Kumar, S., Patil, A. and Hiremath, P.S.,2021. Behavioral biometric data analysis for gender classification using feature fusion and machine learning. Frontiers in Robotics and AI, 8, p.685966.

[8]. Gahmousse, A., Gattal, A., Djeddi, C. and Siddiqi, I., 2020, October. Handwriting based personality identification using textural features. In 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI) (pp. 1-6). IEEE.

[9]. Ghosh, S., Shivakumara, P., Roy, P., Pal, U. and Lu, T., 2020. Graphology based handwritten character analysis for human behaviour identification. CAAI Transactions on Intelligence Technology, 5(1), pp.55-65.

[10]. Sudana, O., Gunaya, I.W. and Putra, I.K.G.D., 2020. Handwriting identification using deep convolutional neural network methods. Telkomnika (Telecommunication Computing Electronics and Control), 18(4), pp.1934-1941.

[11]. Ayzeren, Y.B., Erbilek, M. and Çelebi, E., 2019.Emotional state prediction from online handwriting and signature biometrics. IEEE Access, 7, pp.164759-164774.

[12]. Akhtar, S., Dipti, M.M., Tinni, T.A., Khan, P., Kabir, R.and Islam, M.R., 2021, February. Analysis on Handwriting Using Pen-Tablet for Identification of Person and Handedness. In 2021 International Conference on Information and Communication Technology for SustainableDevelopment (ICICT4SD) (pp. 120-124). IEEE.

[13]. Fatimah, S.H., Djamal, E.C., Ilyas, R. and Renaldi, F.,2019, November. Personality features identification fromhandwriting using convolutional neural networks. In 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)(pp. 119-124). IEEE

[14]. Gattal, A., Djeddi, C., Siddiqi, I. and Chibani, Y., 2018. Gender classification from offline multi-script handwriting images using oriented basic image features (oBIFs). Expert Systems with Applications, 99, pp.155-167.

[15]. Illouz, E. and Netanyahu, N.S., 2018, October.Handwriting-based gender classification using end-to-enddeep neural networks. In the International Conference on Artificial Neural Networks (pp. 613-621). Springer, Cham

Sudeshna S, et. al. "Nationality Identification using Handwriting." *IOSR Journal of Computer Engineering (IOSR-JCE)*, 25(2), 2023, pp. 26-31.