Prognosis of Gender Using Machine Learning Supervised Algorithms

Kartik Pujari, Shruti Rathore
Computer Science and Engineering Department
ABESEC, Ghaziabad.
Miss. Ayushi Agarwal,
Assistant Professor, CSE, ABESEC

Abstract-

This paper aims to predict different traits like; gender, age, and mood recognition using handwriting recognition. Our research work was performed on two published manuscript databases, the IAM dataset and ICDAR dataset containing English texts. Our solution is based on Naive Bayes, KNN and SVM algorithms**.** This is because these models show a better ability to extract good features. This project will also work on machine learning (ML) algorithms but at a later stage, it will be shifted and developed via deep learning (DL) techniques.

Keyword

ML(Machine Learning), DL(Deep Learning), Naive Bayes, KNN (k-nearest neighbor algorithm), SVM(Support Vector Machine), HWR(Handwriting Recognition).

1. Introduction

In this era of fast-growing Technology we have a lot of difficulty in security systems. Nowadays we can see that it has been compromised and lots of technologies are being used for security purposes such as retina scanning, fingerprint unlocking, and even heartbeat detection. These security systems prove that every person is different in some way or another and thus with the help of HWR (handwriting recognition) we can predict some basic traits such as gender, age, and mental state of a person. These traits can help us identify and narrow down a certain person or group of people.

1.1. Motivation

Finding the fundamental human characteristics is difficult to accomplish. Various methods had been used in the past to use HWR to predict traits such as gender, age, handedness, behavior, and many more. One major trait missing in the research work was “mental state prediction”.

1.2. Scope

- Forensics biometrics[6,7] – Helps in the prediction of gender and psychological condition of the person. With the help of HWR, we can determine certain traits and characteristics about the person involved such as depression, temper, anxiety issues, and ill mental state. Handwriting classification helps investigators focus on a certain category of writers.
- Ancient manuscript reading[8] - Ancient manuscript conflicts are being resolved through handwriting recognition, in which the true author of the document is determined using specific characteristics of the author's handwriting. This helps prevent exaggerated claims of manuscripts or handwriting.

• Putting in the mathematical equations by simple handwriting[8] - Handwriting recognition is expected to achieve the goal of automatic online conversion of handwritten math equations on the typed form or student notes. Apps like “photo math” provide such a service through which we can solve complex mathematical equations in a fraction of a second just by scanning them in raw handwritten form.

2. Literature Review:

Table1: Advantages and Drawbacks of various research paper

Name of research paper	**Author’s Name**	**Year**	**Advantages**	**Drawbacks**
Automatic prediction of age, gender, and nationality in offline handwriting	Somaya Al Maadeed and Abdelaali Hassaine	2014	● Pictorial and geometric representation for the classification of age range and gender of HWR.[2]	● Only efficient for English and Arabic datasets.
A Review Paper on General Concepts of “Artificial Intelligence and Machine Learning”	Mahind Rupali, Patil Amit	2017	● Easily identifies trends and patterns.[1] ● Handling multidimensional and multi-variety data.[1]	● High error-susceptibility ● Requires massive data sets.
Neural network model of Artificial Intelligence for HWR	Kulik S.D	2015	● In the paper, the issue of handwritten symbol identification has been looked into.[3] ● This paper introduces a new artificial intelligence neural network model for HWR.[3]	● More gender-determining classes used than required. ● Research on ANN is in the development stage.

Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks	Ángel Morera, Ángel Sánchez, José Francisco Vélez, and Ana Belén Moreno	2017	• This paper has addressed three traits: gender, handedness, and the combined "gender-and-handedness" classification.[4] • CNN had proven better capabilities to extract relevant handwriting.[4]	• Does not mention the domain of recognition of the mental state.
A full English sentence database for off-line handwriting recognition	U.-V. Marti and H. Bunke	2013	• Mentions about large collections of texts, called corpora. These corpora have different appearances and contents.[5]	• Smaller dataset than required. • Doesn't include variety in traits. • Doesn't mention accuracy, recall, f-score, and precision calculation.

Automatic prediction of age, gender, and nationality in offline handwriting

In this Research paper, different categories for classifying manuscripts are discussed. In forensics, handwriting classification helps investigators focus on specific categories of authors. However, little research has been done in this area. The classification of handwritten characters into demographic categories is usually performed in her two steps of feature extraction and classification. The advantage that we observed in this paper is the representation for the classification of age range and gender of HWR in Pictorial and geometric form. And the thing that we find missing is that it is only limited to the English and Arabic datasets.

A Review Paper on General Concepts of "Artificial Intelligence and Machine Learning"

The fundamental ideas of machine learning and artificial intelligence are covered in this research paper. New and highly developed technologies are a divine blessing. Artificial intelligence and machine learning

are two examples of adaptive inventions that have been made for the purpose of decreasing human labor and ensuring a bright future. Even though many early assumptions were incorrect, a new era of faultless technology and superior science is already booming.

Neural network model of Artificial Intelligence for HWR

In the paper, the issue of handwritten symbol identification has been looked into. This paper's major goal is to introduce a new artificial intelligence (AI) neural network model (NNM) for handwriting recognition. Only Russian Cyrillic capital and small letters are used in the handwriting. The first neural network and second neural network are the two subsystems that make up the recognition system. The neural networks were employed to identify handwriting and, in particular, to determine if a writer was male or female. Successful use of the technique (neural network technologies) has been made in the design of a biometric system for the Automated Factographic Information Retrieval System (AFIRS). The AFIRS handwriting recognition neural network algorithm was created.

Gender and Handedness Prediction from Offline Handwriting Using Convolutional Neural Networks

This article offers an experimental analysis of the effectiveness of deep neural networks to three automated demographic issues, including gender, handedness, and a combination of gender and handedness. Applications in fields like forensic biometrics that use demographic handwriting-based classification problems, such as gender and handedness categorizations, are particularly targeted in this. They used the IAM dataset (English texts) and the KHATT dataset (Arabic texts) for the research. Furthermore, the proposed solution was developed using unique network configurations for the three demographic problems considered. This has the advantage of simplifying the design complexity and debugging of these deep architectures when dealing with related hand-written problems.

A full English sentence database for off-line handwriting recognition

In this research paper, along with a few preprocessing and text segmentation techniques, it introduces a novel database for off-line handwriting recognition. The database is founded on
Corpus Lancaster-Oslo/Bergen (LOB). This corpus consists of a collection of writings that were used to create forms that were filled out by the persons manually. The database currently (December 1998) has 556 forms created by roughly 250 different authors. There are whole English sentences in the database. It can act as the foundation for many different handwriting recognition tasks. However, methods of recognition that go beyond lexical knowledge is the main topic of attention. This information can be obtained from outside sources or automatically extracted from the corpus.

3. Data set used:

ICDAR 2013: **International Conference on Document Analysis and Recognition** (ICDAR) is a conference that is held twice a year. This Dataset is considered to be one of the most prominent datasets in HWR (HandWriting Recognition). The data set we collected had 600 handwriting samples. These test cases were part of the ICDAR 2015 edition. It includes English and Arabic samples. In this paper, we will be taking English data into consideration.

IAM: IAM online handwriting database (IAM-OnDB) is a handwriting database for the English language that is used to train and test data. This database was first introduced in ICDAR 1999.

Dataset	Training	Testing	Training Answers
ICDAR 2013	500	205	182
IAM	300	123	100

Figure.1 Datasets used for Testing and Training.

4. Results:

```
Classifier LR, Accuracy: 0.738, AUC: 0.801, Logloss: 0.578
Classifier RF, Accuracy: 0.713, AUC: 0.831, Logloss: 0.558
Classifier LSVM, Accuracy: 0.750, AUC: 0.793, Logloss: 0.584
Classifier SVM, Accuracy: 0.652, AUC: 0.753, Logloss: 0.624
Classifier GB, Accuracy: 0.750, AUC: 0.804, Logloss: 0.768
Classifier ES, Accuracy: 0.726, AUC: 0.813, Logloss: 0.547
>>>
```

Figure.2 Results obtained after testing the data.

By this, we can conclude that the highest accuracy is achieved via the Gradient Boosting (GB) algorithm (75%) and LSVM algorithm (75%). The SVM algorithm provides the least accuracy (65.2). The AUC is highest via the RF algorithm (83.1%) and lowest via SVM (75.3%). Talking about Logloss, the highest logloss is in GB (76.8%) whereas the least is in the RF algorithm (55.8%).

5. References

[1] Rupali, M., & Amit, P. (2017). A review paper on general concepts of artificial intelligence and machine learning. *International Advanced Research Journal in Science, Engineering and Technology*, *4*(Special Issue 4), 79-82.

[2] Al Maadeed, S., & Hassaine, A. (2014). Automatic prediction of age, gender, and nationality in offline handwriting. *EURASIP Journal on Image and Video Processing*, *2014*(1), 1-10.

[3] Morera, Á., Sánchez, Á., Vélez, J. F., & Moreno, A. B. (2018). Gender and handedness prediction from offline handwriting using convolutional neural networks. *Complexity*, *2018*.

[4] Kulik, S. D. (2015). Neural network model of artificial intelligence for handwriting recognition. *Journal of Theoretical & Applied Information Technology*, *73*(2).

[5] Marti, U. V., & Bunke, H. (1999, September). A full English sentence database for off-line handwriting recognition. In *Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR'99 (Cat. No. PR00318)* (pp. 705-708). IEEE.

[6] J. Chapran, "Biometric writer identification: Feature analysis and classification," International Journal of Pattern Recognition and Artificial Intelligence, vol. 20, no. 4, pp. 483–503, 2006.

[7] L. Ballard, D. Lopresti, and F. Monrose, “Evaluating the security of handwriting biometrics,” in Proceedings of the Proc, pp. 461– 466, 2006.

[8] Singh, S. Scope of Handwriting Recognition in Indic Scripts.
[9] Agarwal, A., Saraswat, M., & Arora, A. Leveraging Handwriting For Gender and Handedness Prediction Using Various Machine Learning Approaches.