Handwritten Signature Verification System using machine Learning Approach- A Review of Literature

Pooja Gaikwad[1], Kashaf Pathan[2], Purva Patil[3], Laxmi Pagare[4], R.P. Dahake[5]

[1,2,3,4] Student, Department of Information Technology, MET's Institute of Engineering, Nashik, INDIA

[5] Professor, Department of Information Technology, MET's Institute of Engineering, Nashik, INDIA

ABSTRACT

This system proposes a feasible solution to verify handwritten signatures using various machine learning approaches. The scope has been scaling down to offline signatures which contains static inputs and outputs. Several classification methods such as Multinomial Naive Bayes Classifier (MNBC), Bernoulli Naive Bayes Classifier (BNBC), Logistic Regression Classifier (LRC), Stochastic Gradient Descent Classifier (SGDC), and Random Forest Classifier (RFC) were implemented to identify the most suitable classifier to verify handwritten signatures. The classifiers were pre-trained and tested using a handwritten signature database available for public use available on the Kaggle website. The best performance was obtained from RFC with and accuracy score of more than 0.6. For average, the framework made has been successful in verifying handwritten signature images provided with an extensive precision level.

Keywords— Offline handwritten signature, classification, algorithms, artificial intelligence, Random forest classifier.

I. INTRODUCTION

Signature has been a distinguishing element for individual identification through years. This area incorporates a background marked by ages which begins from the Roman Empire. The legalization of the handwritten signature has been declared by the British government in the 19th century. Signatures have been used for automatic clearing of cheques in the banking industry as well as confirming the legality of any document regarding properties, real estate and agreements. Despite an expanding number of electronic choices to paper cheques and the other materials which are generally authorized by the handwritten signature, fraud detection has been a continuous issue in each area where the handwritten signatures are used. Since banks pay less attention to verifying signatures on cheques and Debit and Shopping cards, a system capable of verifying whether a signature is a forgery or not will prove beneficial. As signature is the primary mechanism for authentication and authorization in legal transactions and documents, the need for an efficient automated solution for signature verification has increased. Therefore, developing a secure and robust platform that automatically authenticates the signatures in few seconds based on the sample dataset of original signatures of the owner is the objective of this system.

II. PROBLEM STATEMENT

The area of handwritten signature identification has been broadly researched in the past few decades and remains as an open research problem. The importance is that the wide usage of the handwritten signatures in verification processes. Even though there are more advanced biometric verification areas such as iris, fingerprint and face recognition, still handwritten signature is mostly accepted as it is used in the legal documentations for decades. Currently handwritten signatures are also used on the back side of the debit electronic cards provided by commercial banks which are used in shopping malls and billing machines. In these points the signature is verified with the signature which will be signed by the owner at the point of transaction. The verification is done manually by the retailer or the operator by investigating and comparing the two versions from the human eye. Therefore, this method of verification depends on the vision capabilities and the perception of the viewer. In addition to that the level of comparison will be decided by the viewer. Other than electronic cards, handwritten signatures are frequently verified in banks when updating accounts and transactions and managing cheques. In most of the times the signatures are compared manually, or they are scanned through an optical scanner and verified. The manual process, as mentioned earlier contains a lot of errors and less accuracy. On the other hand, optical scanners only compare the versions of images taken from the original signature and the testing signature as image pixels and it only checks whether the test image is a perfect reflection of the original one.

A handwritten signature can be varied with time when considering a specific owner. This may be because of aging where the pen pressure, timing and stability of the hand movement could be different. Also, when an owner gets familiar with a signature with the progression of time, it tends to be signed in a shorter time and with the experience details of the signature could be evolved. Due to these reasons a signature may vary with the time where it may not look as same as it was when the original sample was taken. Therefore, optical scanners are not an effective solution to verify handwritten signatures with their dynamic behavior as the scanning process checks for an exact match of the original signature. There are some challenges in Signature Verification - (i) Most of the dynamic information in the signature is lost and (ii) Low quantity of available signature samples versus high number of extracted features.

This study aims to find a feasible solution to verify handwritten signatures with a variation. This research will be focused on cloud-based signature verification, characterized by the usage of static dataset of signatures, where the motive is to discriminate whether a given signature is genuine which means produced by the claimed individual, or a forgery which means produced by an impostor.

III. LITERATURE SURVEY

Igor V. Anikin and Ellina S. Anisimova in “Handwritten signature recognition method based on fuzzy logic” [1], In this paper author suggest a new method for handwritten signature recognition based on fuzzy features of the curvature. In the work authors suggested a method for recognition of fuzzy 2D primitives via a technology of soft computing.

Vijay More et al., [2] observed that segmentation accuracy of Devanagari text characters is completely depends on accurately segmented lines and words from handwritten documents. Author identified many issues and challenges for segmenting lines and words from these handwritten documents. Global threshold and Otsu’s optimum threshold methods were experimented by author and found 85.12% segmentation accuracy.

Yawalkar and M. U. Kharat [3] designed and implemented a segmentation technique for segmentation of multi-connected handwritten Devnagri compound characters to extract the features and recognize the hand-written Devnagri compound characters.

A handwritten signature identification system and the techniques used to solve this problem can be divided into two classes online and Off-line. On-line technique uses an electronic tablet and a stylus connected to a computer system to collect information about a handwritten signature and takes dynamic information like pressure, speed of writing etc. for identification purpose. Offline handwritten signature identification involves less electronic control and uses static signature images captured by scanner or camera. An offline handwritten signature verification system uses features extracted from captured signature image. The features used for offline signature verification are much simpler way. In this only the pixel image needs to be evaluated. But the off-line systems are difficult to design and train to achieve high accuracy as many desirable characteristics such as the order of strokes, the velocity and other dynamic information are not available in the off-line case. The verification process must completely depend on the features that can be extracted from the trace of the static signature images only.

In the field of machine learning, classification is considered an instance of supervised learning, i.e. learning technique where a training data set of correctly identified samples is available. The corresponding unsupervised procedure is known as clustering techniques and involves clustering data into categories based on some measure of inherent similarity or distance.

There are different classifiers that can be used for this application and the followings are the alternatives that was used in this study.

A. **Multinomial Naive Bayes Classifier** - MultinomialNB implements the naive Bayes algorithm for multinomial distributed data, and is one of the two classic naive Bayes variants used in text classification (where the data are typically represented as word vector counts, although tf-idf vectors are also known to work well in practice). If a given class and feature value never occur together in the training data, then the frequency-based probability estimate will be zero. This is problematic because it will wipe out all information in the other probabilities when they are multiplied. [5]
B. **Bernoulli Naive Bayes Classifier - BernoulliNB** implements the naive Bayes training and classification algorithms for data that is distributed according to multivariate Bernoulli distributions; i.e., there may be multiple features but each one is assumed to be a binaryvalued (Bernoulli, Boolean) variable. Therefore, this class requires samples to be represented as binary-valued feature vectors; if handed any other kind of data, a BernoulliNB instance may binarize its input (depending on the binarize parameter).
C. **Logistic Regression Classifier** - In the multiclass case, the training algorithm uses the onevs-rest (OvR) scheme if the 'multi_class' option is set to 'ovr' and uses the cross-entropy loss, if the 'multi_class' option is set to 'multinomial'. (Currently the 'multinomial' option is supported only by the 'lbfgs' and 'newton-cg' solvers) [6]. This class implements regularized logistic regression using the liblinear library, newton-cg and lbfgs solvers. It can handle both dense and sparse input. Use C-ordered arrays or CSR matrices containing 64-bit floats for optimal performance; any other input format will be converted (and copied).
D. **Random Forests** - Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set. [7]

IV. PROPOSED ARCHITECTURE

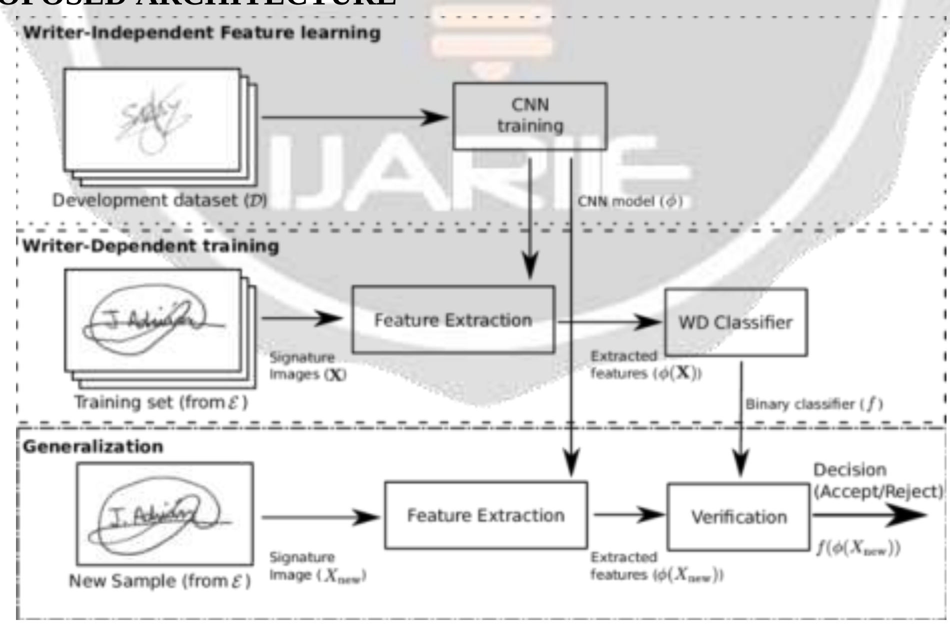

Figure 1 System Architecture

Fig. 1 represents the architecture of proposed handwritten verification system. proposed handwritten signature identification system which identify the authenticity of given signature of a person. The design of a handwriting verification system is divided into two phases: Training phase and Testing phase. A training phase consist of four major steps 1) extraction of a signature image from a dataset 2) pre-processing of image 3) Feature extraction 4) Neural network training. A testing phase consists of five important steps 1) Retrieval of a signature to be tested from a dataset 2) Image pre-processing- Signature preprocessing is a necessary step to improve the accuracy of Feature extraction and Classification and to reduce their computational needs. The purpose of preprocessing phase is to make signatures standard and ready for feature extraction. 3) Feature extraction – The success of a signature verification system greatly depends on Feature extraction. An ideal feature extraction technique extracts a minimal feature set that maximizes interpersonal distance between signature examples of various persons while minimizing intrapersonal distance for those belonging to the same person 4) Application of extracted features to a trained neural network 5) Testing output generated from a neural network.

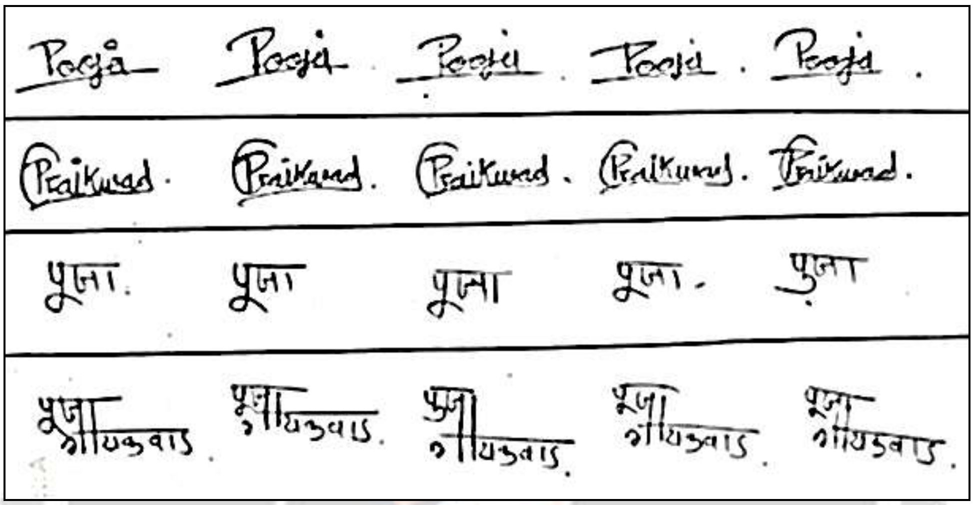

Figure 2 Samples in Latin and Devanagari

IV. CONCLUSION

Verification of handwritten signatures can be done with the use of Random Forest Classifier or the Stochastic Gradient Descent Classifier. This system supports higher variation of signatures and is more flexible. The accuracy level can be controlled by the cut off probability parameter which can be changed by the user according to their needs. In addition to that this system enables the user to do the image processing and classification together in one application. This system can be used as an integrated tool for different domains such as the internal system of a bank or an inventory and sales management system of a retail shop. Higher the number of samples in the training set, higher the accuracy in signature verification.

On the other hand, since the system is flexible and supports variations, forgeries might be promoted compared to the other methods. As signatures varies from time to time and can be causes of frauds since skilled forgeries can be made, there are higher error rates than other traits. Always the inputs of the system are affected by the physical and emotional state of the owner of the signature. In addition to that this system will contain a large temporal variation.

V. REFERENCES

[1] Handwritten signature recognition method based on fuzzy logic Igor V. Anikin;Ellina S. Anisimova 2016 Dynamics of Systems, Mechanisms and Machines (Dynamics) Year: 2016 | Publisher: IEEE

[2] Vijay More, Madan Kharat, Shyamrao Gumaste, "Segmentation of Devanagari Handwritten Text Using Thresholding Approach", International Journal of Scientific & Technology Research (IJSTR), Volume 9, Issue 3, 2277-8616, March 2020, pp.6594-6605.

[3] Prashant Yawalkar, Dr. M. U. Kharat, Dr. S. V. Gumaste, "Segmentation of Multiple Touching Handwritten Devnagari Compound Characters: Image Segmentation

[4] for Feature Extraction", Feature Dimension Reduction for Content-Based Image Identification, IGI Global, A volume in the Advances in Multimedia Kumar and K. Bhatia, "A robust offline handwritten signature

verification system using writer independent approach," 2017 3rd International Conference on Advances in Computing,Communication & Automation (ICACCA) (Fall), 2017.

[5] Ertam and G. Aydin, "Data classification with deep learning using Tensorflow," 2017 International Conference on Computer Science and Engineering (UBMK), 2017.

[6] S. Choi and K. Lee, "A CUDA-based implementation of convolutional neural network," 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), 2017.

[7] L. N. Harfiya, A. W. Widodo, and R. C. Wihandika, "Offline signature verification based on pyramid histogram of oriented gradient features," 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), 2017.

[8] M. S. Chowdhury and M. S. Rahman, "Towards optimal shallow ANN for recognizing isolated handwritten Bengali numerals," 2016 9th International Conference on Electrical and Computer Engineering (ICECE), 2016