An Approach for Hand Written Recognition using Bayesian Network

Rohini A[1], Richa Choudhary[2], Tanupriya Choudhury[3], Sachi N. Mohanty[4], Hitesh K. Sharma[5]

[1] *Anil Neerukonda Institute of Technology and Sciences, Vishakapatnam, AndhraPradesh,531162, India.*
[2]*University of Petroleum and Energy Studies (UPES),Dehradun, 248007,Uttarakhand, India.*
[3,5] *University of Petroleum and Energy Studies (UPES),Dehradun, 248007,Uttarakhand, India.*
[4] *Singidunum University, Serbia and VIT-AP University, Amaravati, Andhra Pradesh, India*.

Abstract

With machine learning, the characters in the digits data set can be recognized. It is a growing concern in pattern recognition. This study analyzes the behaviors of handwritten characters by using the quality of features in handwritten such as font, size, styles of writing, and symbols to make patterns. A process for identifying characters using algorithms. It is difficult to learn algorithms for traditional human writing recognition. The proposed approach extracts the spatial information and applies the bilinear fusion to the flow of patterns. In binary image processing, the Bayesian network is used for classifying the contents and mathematical methods are used to determine patterns in terms of digits. This approach yielded good results.

Keywords

Handwritten Recognition, Bayesian Network Theory, Digits recognition, Bayes Probability.

1. Introduction

It is a traditional approach in machine learning techniques to recognize images and detect them. The research draws the groups from the digits dataset. Images of scanned documents were normalized in size after they were scanned and taken from scanned documents. Using deep learning and machine learning, handwritten character recognition has been used to analyze reading bank check forms, reading postal addresses, and so forth. The handwritten digit recognition has been recognized by the human hand-written digits in the topic of the boundless area of research in the emerging field of deep learning techniques. It's taken from papers, images, and touch screens, etc. The feasibility of digits has been treated as supervised learning types of machine learning and significant sources have been retrieved by user digits to understand and analyze the image recognition. The Quality of lines, space between the words, consistency of sizes, Connectivity of strokes, Pen pressure, are to improve and identify the recognition of handwritten digits. The collected data in terms of user attributes and typed characters. The diversities of writing types which are space between the characters, diminish of letters are challenges to identify the digits in the dataset. The co-occurrence of the digits has the evidential datasets of the image structure. Strokes are inherently strong connections for dealing or inferring data to join the digits in the data set. The complexity of data to recognize the images has more difficult to identify the target node. A large enormous of data has been taken from the MNIST database to predict the digits given an image. The recognition of image is carried out three steps. (a) Input of the segmentation into hypothetical symbols. It has been recognized by the symbol classifier and determined the structural expressions. It takes into different forms of neural .Behind

ACI'22: Workshop on Advances in Computation Intelligence, its Concepts & Applications at ISIC 2022, May 17-19, Savannah, United States
EMAIL: rohinaruna@gmail.com (A. 1); richachoudhary.86@gmail.com(A. 2); tanupriya1986@gmail.com (A. 3); sachinandan09@gmail.com(A. 4); hkshitesh@gmail.com(A. 5)
ORCID: 0000-0001-5809-7317 (A. 1); 0000-0003-3277-9920 (A. 2); 0000-0002-9826-2759 (A. 3); 0000-0002-4939-0797 (A. 4); 0000-0001-6816-0324 (A. 5)

the recognization of patterns image processing has carry over to analyses the strokes and strings. The study has been elaborated the approximation of mathematical equations for recognizing the strokes in the handwritten digits. The digit has increased in the image; it's become more complicated to identify the input stroke to symbols. That symbols are said to be hypothetical symbols. It has been recognized by the classifier. The symbol structures are recognized by the expression of structures and it has been analyzed by the parsing algorithm. It provides the tool for analyzing the structure and coherent approaches are used for recognition and identifying the images by using a mathematical equation. The pattern recognition tool is used to analyze the information from raw handwritten images or digits. There are different styles of handwriting in different communities. Images are not in sharp the redundancy of images were removed. This approach has to be implementing with the SVM recognition system to perform by the NIST SD19 data set (https://www.nist.gov/srd/nist-special-database-19). Patterns are represented mathematically to understand the dynamics tools for communication and information. For instance, a handwritten character set drives the different styles of writing. Hence, if topological properties of the handwritten dataset are connected, which signifies the redundancy of data, and sharp recognition. The structural dimension provides to benefit the terms of accessing the large structured images.

2. Materials and Methods

The efforts are carried out by pattern expressions and recognition of characters. The images were extracted from the user interaction data along with the properties of communication. The Extracted data has been analyzed and tested by R-Programming for statistical analysis. It is used to provide functions and data types for identifying the patterns. A supervised Learning algorithm has easy to identify the strokes and structure, size, and intensity of images in the data set. Handwritten mathematical equations have been tested in the proposed study and compared with the existing algorithms. The dataset consists of 5000 training sets and 1000 for testing data set which helps find the patterns of recognition. Images were accommodate and tested the Machine Learning algorithms of Support Vector Machine, K-Nearest Neighbor and Convolution Neural Network.

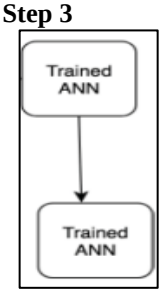
Image Acquisition: The image has been captured from the scanner. It has numbers, symbols, characters and a special set of characters has been collected and put into the sequential form of a database.

Preprocessing: Given image has converted into a greyscale image with a threshold value of 0.5. It removes blur spots from images and it inverts and reshapes the images.

Step 1

Step 2

Step 3



In the handwritten recognition to improve the features has been computed by the input images. The process of detection and correlation is the challenging problem in the preprocessing images. Generally categorized the process into removal of noise, normalization, and smoothing the images in the input images.

The input image has been formatted up to the saturation level of gray scale image. Then it has been turned into binary format. By the two categories binarization method has taken place. (1) Global threshold (2) Local Threshold.

In the algorithm of Global Threshold: The study has taken a single threshold value for the overall images based on the uniformity and measuring the shapes in the image.

Local Threshold: the angle has been determined by the threshold values for pixel using their spatial information. After binarization pixel densities and angle of the text is chosen in the text image to realize the actual angle.

Segmentation: The recognition of the process has been analyzed by the segmentation. The size of each digit and gap between the digits are unknown. To fill the gap by the digit segmentation algorithm to remove the noisy images and connect the components in the hand written digit. The aim of the paper is to crop the correct segmentation for handwritten recognition by using single touch and multiple touches of strings.

Functional Analysis is used to separate the features of digits.Input images are resized into 5*7 pixels towards the training data.

Feature Extraction: By the structural extraction the morphological features of edges, regions, and curves has been analyzed for indexing and labeling the data to help the classification of handwritten digits. It extracts the information from the input image. The parametric features were area, centroid, density, line segment also extracted to find the weight and center of location in the box area.

Classification: By using soft computing techniques of multilayer perceptron is arranged in layers which are input layer, hidden layer, and outer layer of the node. It gives the classification performance of extracting the features. Step 1,2,3 describes how from the input images , the binarization and features and testing's are processed.

3. Bayesian Network Approach

BN is a system to minimize the classification error and plays a role in the prior probability of information. It is a statical approach to quantify the various decision.

The conditional probability and prior probability has been computed as:

$$P\left(\frac{m_n}{x}\right) = {P(x/m_n) * P\ (m_n)}/{P(x)} \quad (1)$$

Where $P(x) = \sum_{j=1}^{k} P(x/m_n) * P\ (m_n)$ (2)

$P(x/m_n) * P\ (m_n)$ -- Likelihood event

$P(x)$ -- Probability of Evidence

Outliers of the decision probability

$P(Error/X) =$ $P(m_1/x)$ if we choose m_1,

$P(m_2/x)$ if we choose $m_{2,}$

Observe x and action αi, the true nature of state is $\alpha_{i.}$ Incur the $\lambda(\alpha i|m_n)$.

R (αi |x) = Conditional Risk.

4. Simulation Results

The implementation results were shown below, the mapping of the model has been implemented in Graphical User Interface components of Java programming by the eclipse tool.

Fig. 1, Shows the Handwritten character has converted into binary, removal, and segmentation. Fig. 2, shows the performance of feature extraction using Bayesian Decision theory.

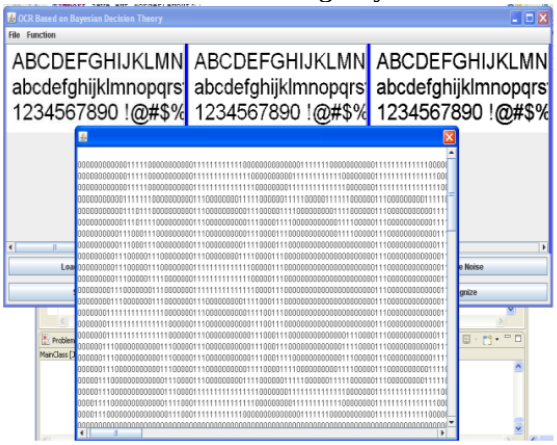


Figure 1: Segmentation of Noise removal and Binary classification

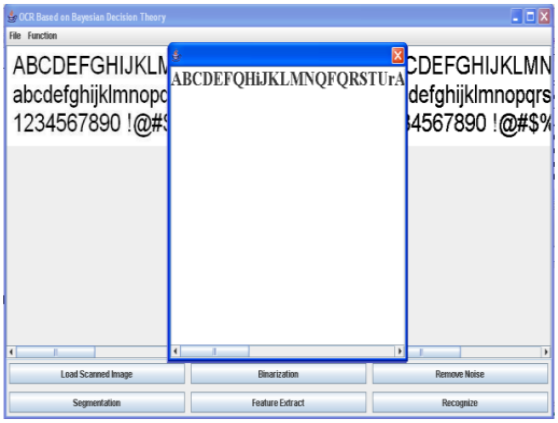


Figure 2: Recognition of Digits using Bayesian Approach

5. Result and Discussion

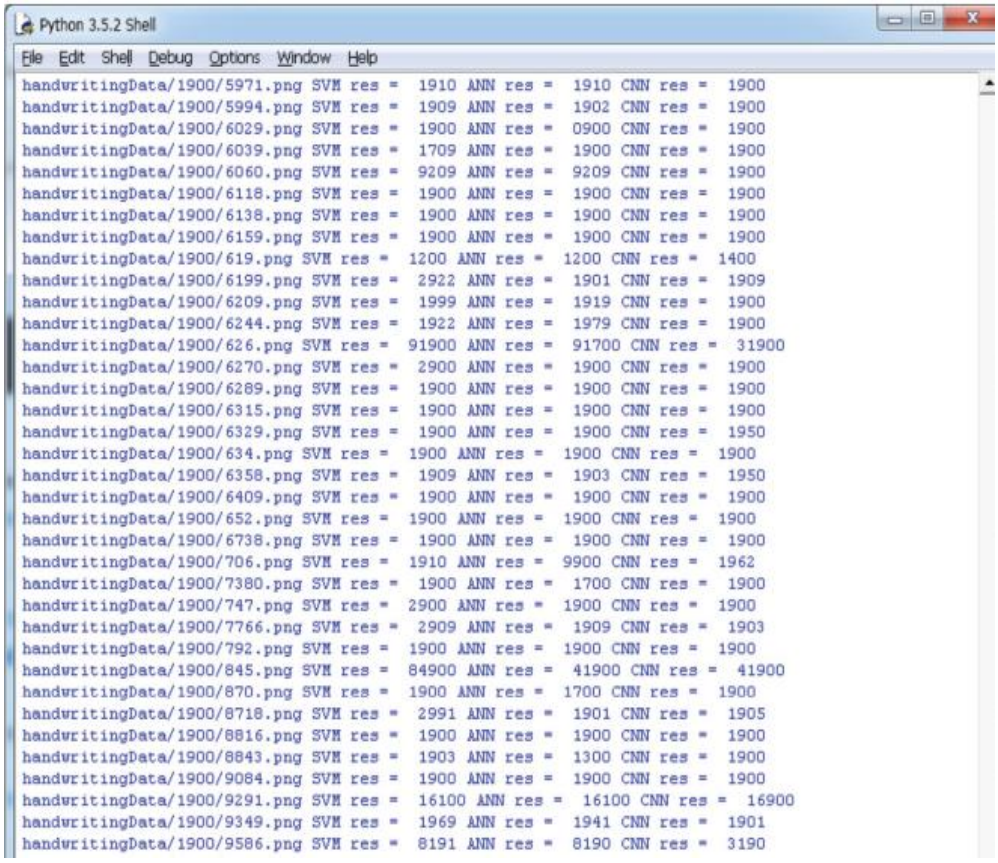


Figure 3: Compare Accuracy of Learning algorithms

Word recognition is the number of words in transcription. we combined the online and offline mode of the system have recognized and measured the rate has compared with four existing algorithms shown in Fig.3. which are KNN, SVM, CNN, and BN, BN has incrementally recognized the rate of recognition as 75.20% and the accuracy as 66.10%. Each output position of KNN recognition rate as 65.90% and the accuracy as 61.4%. SVM as 73.4% and accuracy as 65%when compared to KNN, SVM has a high accuracy rate. Eventually, the CNN recognition rate has 73.8% and accuracy was 65.3%.

6. Conclusion

By using the appropriate parameter parameters of feature and quality of recognition, the proposed work of handwritten character recognition fits into the input image of the meta-database. Based on the Bayesian network approach, we have classified effectively, decreased errors, and increased accuracy in character recognition.

7. Future Work

To propose new classification models to improve the performance of segmentation and reduce the complexity of algorithm to enhance faster computation.

8. References

[1] approached the content-based classification used by the techniques of K-Nearest Neighbour, Decision tree, and Support Vector Machine for analyzing the classification techniques and compared the results. SVM has not been supported and many outliers are identified in the dataset. It becomes a core problem identifying the image classification and clustering the data. [2]proposed the dimensionality reduction techniques used for handwritten recognition with different images in the document and it has been analyzed by the parameters of extending the radius of complex background with high intensities of detersive images. The pre-processing has included in the stretches of images, normalization, binarization, segmentation, and morphological techniques. When compared the accuracy of an image has very less and up to the mark has been identified. [3]analyzed the performance of the neural network with geometric and gradient features. It has been efficiently observed through the Levenberg algorithm. [4] proposed the nearest neighbor method to classify the images with different wavelets. [5] compared the efficiency with the existing algorithms of Hierarchical, Anamoly detection, Apriori Algorithm, and Principal Component Analysis(PCA). PCA has yielded a good classification and efficiency of data. [7][8] implemented the character recognition using deep learning and machine learning techniques. It has been used by the Random Forest, Neural network, and Support Vector Machine.

[1] Mahmoud M. Abu Ghosh; Ashraf Y. Maghari, “A Comparative Study on Handwriting Digit Recognition Using Neural Networks”, IEEE, 2017.

[2] H. Du, P. Li, H. Zhou, W. Gong, G. Luo, and P. Yang, "WordRecorder: Accurate Acoustic-based Handwriting Recognition Using Deep Learning, "IEEE INFOCOM 2018 - IEEE Conference on Computer Communications, 2018, pp. 1448-1456, DOI: 10.1109/INFOCOM.2018.8486285.

[3] V. V. Mainkar, J. A. Katkar, A. B. Udate, and P. R. Pednekar, "Handwritten Character Recognition to Obtain Editable Text," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020, pp. 599-602, DOI: 10.1109/ICESC48915.2020.9155786.

[4] J. Dalal and S. Daiya, "Image processing based optical character recognition using Matlab", *International journal of engineering sciences & research technology*, vol. 7, no. 5, pp. 406-411, 2018.

[5] Anuj Dutt, AashiDutt,” Handwritten Digit Recognition Using Deep Learning”, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 6, Issue 7, July 2017, ISSN: 2278 – 1323.

[6] M. Zanoni, “Data mining techniques for design pattern detection,” Ph.D. dissertation, Universita degli Studi di Milano-Bicocca, 2012.

[7] Handwritten Digit Recognition Using Machine Learning https://medium.com/@himanshubeniwal/handwritten-digitrecognition-using-machine-learning-ad30562a9b64”.

[8] J. Niere, W. Schafer, J. Wadsack, L. Wendehals, and J. Welsh, "Towards pattern-based design recovery," in Proceedings of The 24th International Conference on Software Engineering, 2002, pp. 338–348.

[9] Darmatasia and Mohamad Ivan Fanany,” Handwriting Recognition on Form Document Using Convolutional Neural Network and Support Vector Machines (CNN-SVM), Machine Learning and Computer Vision Laboratory Faculty of Computer Science, University of Indonesia.

[10] M. Elleuch, R. Maalej, and M. Kherallah, “A New Design Based - SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition,” Procedia - Procedia Comput. Sci., vol. 80, pp. 1712–1723, 2016.

[11] Tina Vaz , Nagaraj Vernekar,”A Survey on Evaluating Handwritten Iterative Mathematical Expressions”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2019.

[12] Malla V, S., Sharma, H.K., Choudhury, T. (2022). Handwritten Digit Recognition with Neural Network. In: Sharma, D.K., Peng, SL., Sharma, R., Zaitsev, D.A. (eds) Micro-Electronics and Telecommunication Engineering . Lecture Notes in Networks and Systems, vol 373. Springer, Singapore. https://doi.org/10.1007/978-981-16-8721-1_51.
[13] M. Mishra, T. Choudhury and T. Sarkar, "Devanagari Handwritten Character Recognition," 2021 IEEE India Council International Subsections Conference (INDISCON), 2021, pp. 1-6, doi: 10.1109/INDISCON53343.2021.9582192.
[14] Chakraborty, P., Jahanapi, S.S., Choudhury, T. (2022). Bangla Handwritten Digit Recognition. In: Tavares, J.M.R.S., Dutta, P., Dutta, S., Samanta, D. (eds) Cyber Intelligence and Information Retrieval. Lecture Notes in Networks and Systems, vol 291. Springer, Singapore. https://doi.org/10.1007/978-981-16-4284-5_14.