International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0025
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 144-149

Handwriting Recognition Implementation: A Machine Learning Approach

Dr. A. Shaji George
Almarai Company, ITSM, EIT Department, Riyadh, Kingdom of Saudi Arabia.
Emails: *drashajigeorge@gmail.com*

Abstract
Handwritten text recognition, also referred to as handwritten character recognition, is a field of study that combines model recognition, computer vision, and artificial intelligence. In order to translate handwritten letters into relevant text and computer commands in real time, handwriting recognition systems use pattern matching. The properties of photographs and touch-screen devices can be acquired, detected, and converted into a machine-readable form by an algorithm that recognizes handwriting. An ensemble of bagged classification trees is one way to accomplish this. A bagged classification tree is an ensemble learning technique that helps to increase the efficiency and accuracy of machine learning algorithms by lowering the variance of a prediction model and addressing bias-variance trade-offs. The standard Kaggle digits dataset from (0-9) was utilised in this study to identify handwritten digits using a bagged classification method. And with an accuracy level of 0.8371, we finally came to a conclusion about the importance of the bagged classification strategy.
Keywords: *Convolution Neural Network, machine learning, Bagged, OCR.*

1. Introduction
Machine learning aims to extract hidden information that is present in the data using knowledge of current data on a certain subject. We can achieve machine learning and predict results for unknown data by using specific mathematical functions and concepts to uncover hidden information. One of the key uses for ML is pattern recognition. Large picture data sets are typically used to recognize patterns. An example of pattern recognition through an image is handwriting recognition. We may teach computers to interpret letters and numbers from any language that are contained in an image by employing such notions. Handwritten characters can be recognized using a variety of techniques. Handwritten digits recognition problems can be solved using machine learning models such as artificial neural networks, convolutional neural networks, decision tree classifiers, and improved chain code histogram features [1]. The practice of identifying numbers written by humans using a machine learning approach is known as "recognizing hand-written numbers"[2]. The fact that a machine can complete a job based on training suggests that the recognition of handwritten digits is a significant research problem [3]. The issues with machine learning methods. Humans are not required to write a line of text in a straight line on white paper, although the text in printed documents is in a straight line.

- Character differentiation and identification are more difficult in cursive writing.
- Unlike printed text, which is always straight, handwritten text can rotate to the right.
- Compared to synthetic data, gathering a good set of tagged data to learn is expensive.
- The attacks differ greatly and are vague from person to person. [1-5]
- One's writing style likewise fluctuates and is unpredictable from one instant to the next.
- Poor quality of the original document or image as a result of ageing.

2. Literature Survey
This section will go over handwritten figure recognition research that have been used in various contexts. The neural network for handwritten figure identification can be used to scan handwritten figures, extract attributes, and create classifiers [5]. The method of handwriting analysis is time-consuming, systematic, and extensive. Understanding how people write their letters or numbers in a way that takes advantage of the distinctive properties of these

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0025
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 144-149

symbols, such as their sizes, forms, and writing styles [6]. The acronym CNN stands for Convolution Neural Network. The word "convolution" means to twist or coil. The human brain is comparable to any neural network. The brain serves as an inspiration while designing neural networks. The major application of CNN is image categorization. Depending on the needs, CNN has numerous layers. Using two separate learning algorithms, Ahmed Mahdi Obaid and his coworkers [9] devised an efficient handwritten text recognition system. In comparison to Resilient Back-propagation algorithm, Scaled Conjugate Gradient algorithm shown superior performance in terms of accuracy and training time.

3. Methodology

The most sought-after skill in machine learning is now object recognition. Face recognition, handwriting recognition, disease diagnosis, and other examples of object recognition are a few. All of these things are possible thanks to a sizable image data set. Both positive and negative information pertaining to that domain will be included in this image data set. It can serve as a foundational feature for the development of new requirements. For instance, unless a newspaper is in braille format, a blind person cannot read it. In this situation, we may train the algorithm to detect the newspaper characters, save them as text, and then translate the text into speech.

This can make regular tasks easier for many blind people. Language translation may be the second use for handwriting recognition. When dealing with a language other than one's own, one can simply take an image of a document and send it to the algorithm for handwriting recognition. The characters in an image can be recognised by this method and converted to text. The text can then be translated into the chosen language. Processing a vast collection of paper documents, such as test scripts, is another use for handwriting recognition. The evaluation of the answer scripts can be done automatically with the aid of AI and hand-write recognition. Handwriting recognition serves as the basis case to be resolved for all of the aforementioned possibilities. One form of optical character recognition is the ability to recognize handwriting (OCR). OCR is the process of identifying text, whether it is printed or written by hand. In OCR, the document is photographed as an image and then converted to the required format, such as a PDF. The character recognition algorithm is then given the file. In some situations, this can significantly reduce the amount of human interaction. Printed character recognition and Hand-written character recognition are the two OCR derivatives. As the name implies, printed character recognition is the identification of characters in an image of a newspaper or other printed document. Characters written by humans or influenced by humans are recognized when they are written in handwriting. Online and offline character recognition are the two categories into which it is classified. [6-10]

3.1 Image source

Offline handwritten character recognition includes this stage. Any digital instrument may be used as the image source. The image is captured by a scanner or camera and moved on to the following stage.

3.2 Pre-processing

Pre-processing is a series of steps that enhances image quality and, as a result, increases image accuracy. The following pre-processing techniques are used for handwritten character recognition.

3.3 Segmentation

Segmentation is a method for removing specific characters from an image. Segmentation comes in two flavours. Both implicit and explicit segmentation are used. In implicit segmentation, the procedure of segmenting words is not used. However, explicit segmentation predicts words by removing each individual character.

3.4 Feature-extraction

The algorithm for recognition begins with this stage, which is one of the most crucial ones. Each character possesses unique qualities. It consists of a set of rules, each of which explains a character feature. In this phase, such traits are extracted.

3.5 Classification

The training would have ended by this point, and the testing of the input data would have begun. The testing data would run through all of the aforementioned steps, and different probabilities are given to the matching rules. The rule with the highest likelihood is chosen, and the associated class-label is given a recognisable character.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0025
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 144-149

3.5.1 Data Sets

We used a Kaggle dataset in this paper. The handwritten digit recognition issue is the main objective of the Kaggle dataset, which is used to assess machine learning models [7]. It has gained the most traction in the effort to put handheld identification devices into practise. 21,000 forms and 21,000 handwritten number tests are included in the Kaggle dataset (0 to 9). Each number is normalised, slanted, and converted into a 28x28 pixel grayscale (0–255) image. The 784 pixels in each image stand in for the number structures. In Figure 1, the Kaggle dataset is displayed. [11-15]

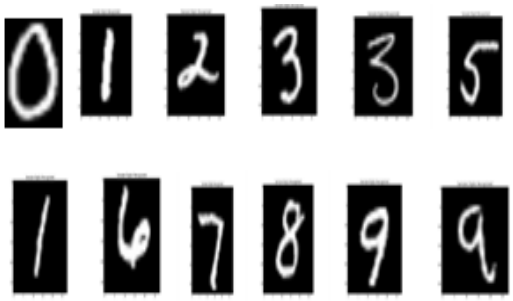

Figure 1 Example of Kaggle Dataset

3.5.2 Algorithm

For handwriting recognition in this paper, we employ the Bagged Classification tree technique. The primary concept behind bagged trees is that instead of relying on one decision tree, we rely on several, which enables us to benefit from the knowledge of numerous models. When evaluating a model's performance, we frequently consider what is known as the bias-variance compromise of our output. The discrepancy has to do with how our model handles minor errors and how such can interfere with it and introduce bias. Effectively, the model makes false assumptions about the connections between variables. It will be wholly inaccurate, regardless of whether the model is accurate or not in terms of direction. A decision tree's significant variance is the main cause for concern. The issue is that even little changes to the data can have a significant impact on the model and forecasts for the future. This is relevant because one of the advantages of bagged trees is that they help reduce variances while maintaining bias[8]. The bagged classification tree's operational phases.

- Divide the data into sets for training and validation at random
- Using Bagged Decision Trees, train and forecast
- Use the Confusion Matrix
- Load Test and Training Data
- Display samples from the Kaggle dataset

3.5.3 Ensemble Learning

A popular machine learning technique known as ensemble learning combines numerous separate models, often known as base models, to create an efficient, ideal prediction model [9].

3.5.4 Bagging in Machine Learning

An ensemble learning technique called bagging, often referred to as Bootstrap aggregating, aids in enhancing the efficiency and precision of machine learning algorithms. It lowers the variance of a prediction model and is used to handle bias-variance trade-offs. Figure 2 illustrates bagging, which is used for decision tree algorithms in both classification and regression models to prevent overfitting of data. Figure 2 shows Bagging in Machine Learning

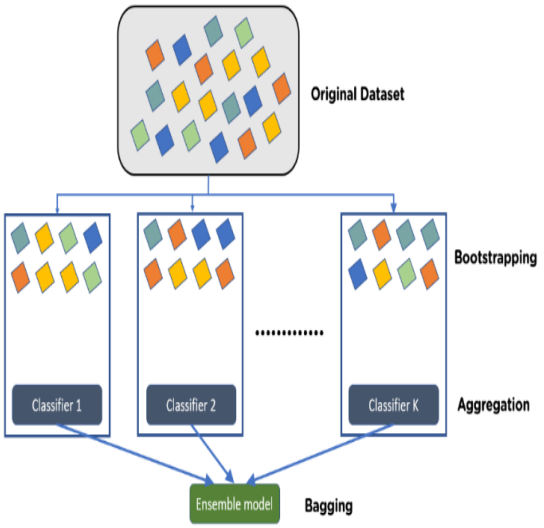

Figure 2 Bagging in Machine Learning

3.5.5 Bootstrapping

Figure 3 illustrates the bootstrapping technique, which involves selecting data samples at random from a population with replacement in order to estimate a population parameter. Figure 3 shows Bootstrapping in Machine Learning.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0025
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 144-149

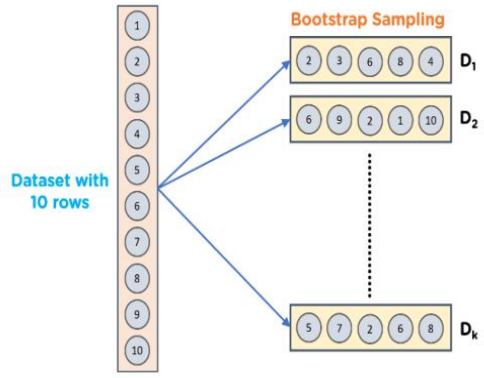

Figure 3 Bootstrapping in Machine Learning

4. Experimental Results

The experiment used the bagged classification tree to train the computer on handwritten digits from the common Kaggle dataset. Nearly 42,000 photos of example handwritten digits are included in this dataset. We used 21,000 of these photos for training and the remaining 21,000 for testing. We used the numbers 1, 2, 3, and 7 to test the bagged classification. Following the machine's training on this dataset, the test results are shown in Table 1. For instance, the handwriting style similarity between 7 and 1 led to the prediction of 7 as 1. Table 1 shows Test Results After Training the Machine On the Dataset.

Table 1 Test Results After Training the Machine On the Dataset

Handwritten Digits images/ Test Data	System Digit Prediction	Expected Prediction
	Correctly Predicted	1
	Correctly Predicted	2
	Correctly Predicted	3
	Correctly Predicted	7

5. Confusion Matrix

The perplexity The package's matrix function is really helpful. to evaluate the effectiveness of a categorization model. Fill the parcel with your forecasts and the actuals, then deliver it. Figure 4 illustrates the accuracy level of this tree, which is 0.8371 when compared to a single decision tree.

- The quadrant cell where both the reference and the forecast are 1 is considered to be true positive. This shows that your prediction about their survival came true.
- False positive #2: You expected a favorable outcome here, but you were off.
- A true negative is when your prediction of the opposite is accurate.
- False negative: When you forecast something negative but are mistaken. Figure 4 shows Confusion Matrix Ensemble Bagged Classification Trees [16-20]

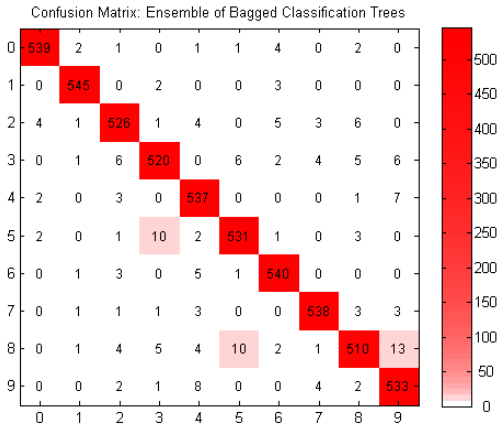

Figure 4 Confusion Matrix Ensemble Bagged Classification Trees

Conclusion

In this study, we use a Bagged decision tree classification model to apply machine learning to recognise handwritten numerals. Using a Kaggle dataset with 41000 rows and 730 columns, the machine's accuracy was trained, and the results are 84.02%. The experiment's findings show that the Bagged classification tree classifier can accurately recognise handwritten numerals. We used the Kaggle repository to store our training and test datasets.

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0025
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 144-149

References

[1]. Salma Shofia Rosyda, and Tito Waluyo Purboyo. "A Review of Various Handwriting Recognition Methods." International Journal of Applied Engineering Research 13.2 (2018): 1155-1164.
[2]. Younus, S. B. S., S. Shajun Nisha, and M. Mohamed Sathik. "Comparative Analysis of Activation Functions in Neural Network for Handwritten Digits." Studies in Indian Place Names 40.71 (2020): 793-799.
[3]. Jagan Mohan Reddy D, A Vishnuvardhan Reddy "Recognition of Handwritten Characters using Deep Convolutional Neural Network".
[4]. Hruday M. "Implementation of Handwritten Character Recognition using Neural Network".
[5]. Kh Tohidul Islam, Ghulam Mujtaba, "Handwritten Digits Recognition with Artificial Neural Network," Proc. of the International Conference on Engineering Technologies and Technopreneurship (ICE2T 2017), Sep 2017.
[6]. Haider A. Alwzwazy, Hayder M. Albehadili, Younes S. Alwan, Naz E. Islam. "Handwritten Digit Recognition Using Convolutional Neural Networks," International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization), vol. 4, no. 2, Feb 2016.
[7]. Vijaya Kumar, A Sirikirshna, Raveendra Babu and Radhika Mani, "Classification and recognition of handwritten digits by using mathematical morphology," Indian Academy of Sciences, 2010.
[8]. Areej Alsaafin, Ashraf Elnagar, "A Minimal Subset of Features Using Feature Selection for Handwritten Digit Recognition," Journal of Intelligent Learning Systems and Applications, vol. 9, pp. 55-68, 2017.
[9]. J. J.Wang, S.G. Hu1, X.T. Zhan, Q.Yu1, Z. Liu, T. P. Chen, Y.Yin, Sumio Hosaka, Y. Liu, "Handwritten-Digit Recognition by Hybrid Convolutional Neural Network based on HfO2 Memristive Spiking-Neuron," Sceintifi Reports, Aug 2018
[10]. Stuti Asthana, Amitkant Pandit, Dinesh Goyal, "A Literature Survey On Better System Efficiency Of Handwritten Numeral Recognition," Suresh Gyan Vihar University International Journal of Environment, Science and Technology, vol. 3, no. 1, Jan 2017.
[11]. Surya Nath RS, and S. Afseena. "Handwritten Character Recognition–A Review." International Journal of Scientific and Research Publications (2015).
[12]. Meenu Mohan, and R. L. Jyothi. "Handwritten Character Recognition: A Comprehensive Review on Geometrical Analysis."
[13]. Yuval Netzer, Tao Wang Adam Coates Alessandro Bissacco Bo Wu Andrew Y. Ng. "Reading digits in natural images with unsupervised feature learning." (2011).
[14]. Polaiah Bojja., Naga Sai Satya Teja Velpuri, Gautham Kumar Pandala,S D Lalitha Rao Sharma Polavarapu "Handwritten Text Recognition using Machine Learning Techniques in Application of NLP".
[15]. Ahmed Mahdi Obaid, IIHazem M. El Bakry, IIIM.A. Eldosuky, IVA.I. Shehab "Handwritten text recognition system based on neural network." Int. J. Adv. Res. Comput. Sci. Technol.(IJARCST) 4.1 (2016): 72-77.
[16]. S. M. Shamim, Mohammad Badrul Alam Miah, Angona Sarker, Masud Rana, Abdullah Al Jobair "Handwritten digit recognition using machine learning algorithms." Global Journal Of Computer Science And Technology (2018).
[17]. P. Shankar Rao, J. Aditya. "Handwriting Recognition – "Offline" Approach".
[18]. Lavanya K, Shaurya Bajaj, Prateek Tank Shashwat Jain, "Handwritten Digit Recognition using Hoeffding tree, Decision tree and Random forests Comparative

International Research Journal on Advanced Engineering and Management
https://goldncloudpublications.com
https://doi.org/10.47392/IRJAEM.2025.0025
e ISSN: 2584-2854
Volume: 03
Issue:02 February 2025
Page No: 144-149

Approach," 2017 International Conference on Computational Intelligence in Data Science(ICCIDS),2017.

[19]. Stuti Asthana1, Amitkant Pandit, Dinesh Goya, "A Literature Survey On Better System Efficiency of Handwritten Numeral Recognition," Suresh Gyan Vihar University International Journal of Environment, Science and Technology, vol. 3, no. 1, pp. 1-5, Jan 2017.

[20]. Pooja Yadav, Nidhika Yadav, "Handwriting Recognition System- A Review," International Journal of Computer Applications, vol. 114, no. 19, pp. 46-40, Mar 2015.