



# Comparative Analysis: Handwritten Character Recognition Via K-Nearest Neighbour And Naive Bayes Algorithms

N. Penchalaiah

*Audisankara College of Engineering and Technology, Gudur, Andhra Pradesh, India*

\*Corresponding author Email: [pench75@gmail.com](mailto:pench75@gmail.com)

**Abstract.** The Study of Handwritten Character Prediction: Comparative Analysis Between K-Nearest Neighbour and Naive Bayes Algorithms This research aims to predict handwritten characters by employing the K-Nearest Neighbour (KNN) algorithm and comparing its feature extraction precision with the Naive Bayes (NB) algorithm. Both algorithms were evaluated using a sample size of  $n = 25$  and iterated 15 times to enhance accuracy and reduce errors, with a statistical power of 80% (G power) and an alpha value of 0.05. The experimental results demonstrated a significant improvement in the accuracy of handwritten character recognition, achieving a significance value of 0.165 through the comparison of NB and KNN algorithms. Comparative analysis revealed that the NB algorithm (97.45) outperformed the KNN algorithm (97.3). These findings highlight the superiority of the NB algorithm in predicting handwritten characters over the KNN algorithm.

**Keywords:** Handwritten character recognition; K-Nearest Neighbour (KNN) algorithm; Naive Bayes (NB) algorithm; Comparative analysis; Feature extraction precision

## 1. Introduction

Handwritten character recognition is the learning method to the computer to recognize and predict the handwritten characters using the machine learning algorithms. The algorithms which are used will convert the input into machine understandable form [1]. The applications i.e., recognizing handwritten characters are used in scanned documents, historical preserved documents, images, bank cheques [2]. As every individual has different handwriting, written by hand character acknowledgment is the learning strategy to the PC to perceive and foresee the transcribed characters utilizing the AI calculations. The calculations which are utilized will change over the contribution to machine justifiable structure [1]. The applications i.e., perceiving written by hand characters are utilized in examined reports, chronicled saved records, pictures, bank checks [2]. As each individual has different penmanship, this transcribed person acknowledgment helps in examining the characters or letters given [3]. Same dataset is utilized for both the calculations to analyze the calculations. By working on the exactness of the models the KNN accomplished 97.3% and NB accomplished 97.4% [4]. In excess of 140 references have been distributed on written by hand character acknowledgment utilizing KNN and NB calculations. Measurable method is utilized for perceiving manually written person acknowledgment which depends on Bayes hypothesis. The method is additionally called a Bayesian channel. The strategy utilized in the proposed work is a successful procedure and the exploratory outcomes acquired by utilizing the method have a high acknowledgment execution [5]. Five sorts [6] of skeleton highlights utilized are crossing focuses, front and back, vertical and even. Water supply which has elements, for example, number of openings, fill-opening thickness, greatest profile distance is utilized as another methodology in written by hand character acknowledgment. By utilizing this new methodology, the KNN calculation acquired 96.94% exactness in perceiving the characters. The proposed framework utilized corner to corner and progress highlights of the pictures as a proposed framework in investigating the transcribed characters [7]. By the new progress and askew technique, the calculation accomplished an exactness of 94.12%. By binarization the dataset pictures will be cleared, and the wanderer pixels of the dataset are removed. The extraction of highlights from the pictures of the dataset are finished by utilizing the strategies called Discrete Wavelet Transform and Discrete Cosine Transform. By utilizing the new methodology, the acquired exactnesses by the calculations are 97.4% and 89.51% [8]. The proposed strategy is preprocessing the information and normalizing the information for written by hand character acknowledgment. With practically no improvement in preprocessing steps the model accomplished a precision of 92.86% [9]. Already our group has a rich involvement with chipping away at different examination projects across numerous disciplines [10-24]. There was a hole in the generally existing frameworks which manage the KNN and NB calculations. Since it very well might be challenging to foresee the person and digit acknowledgment precisely and unequivocally. It is critical to add more qualities to the dataset and prepared dataset to foresee precisely. As a creator, the AI based value expectation arrangement has been done [25]. This exploration point is to foresee transcribed person acknowledgment with better precision utilizing KNN and NB calculations. This handwritten character recognition helps in analysing the characters or letters given [3]. Same dataset is used for both the algorithms to compare the algorithms. By improving the accuracy of the models, the KNN achieved 97.3% and NB achieved 97.4% [4]. More than 140 citations have been published on handwritten character recognition using KNN and NB algorithms. Statistical technique is used for recognizing handwritten character recognition which is based on Bayes theorem. The technique is also called a Bayesian filter. The technique used in the proposed work is an effective technique and the experimental results obtained by using the technique have a high recognition performance [5]. Five types [6] of skeleton features used are crossing points, front and back, vertical and horizontal. Water reservoir which has features such as number of holes, fill-hole density, maximum profile distance is used as a new

approach in handwritten character recognition. By using this new approach, the KNN algorithm obtained 96.94% accuracy in recognizing the characters. The proposed system used diagonal and transition features of the images as a proposed system in analyzing the handwritten characters [7]. By the new transition and diagonal method, the algorithm achieved an accuracy of 94.12%. By binarization the dataset images will be cleared and the stray pixels of the dataset are removed. The extraction of features from the images of the dataset are done by using the techniques called Discrete Wavelet Transform and Discrete Cosine Transform. By using the new approach, the obtained accuracies by the algorithms are 97.4% and 89.51% [8]. The proposed method is preprocessing the data and normalizing the data for handwritten character recognition. Without any improvement in preprocessing steps the model achieved an accuracy of 92.86% [9]. Previously our team has a rich experience in working on various research projects across multiple disciplines [10-11]. There was a gap in the already existing systems which deal with the KNN and NB algorithms. Since it may be difficult to predict the character and digit recognition accurately and precisely. It is important to add more values to the dataset and trained dataset to predict accurately. As an author, the machine learning based price prediction solution has been carried out. This research aim is to predict handwritten character recognition with better accuracy using KNN and NB algorithms.

**2. Materials and Methods**

The review on both the gatherings was led in Saveetha School of Engineering. The review is in examination of two gatherings: one gathering is NB and the other gathering is KNN. The calculations were led on the example size of 25 and iterated for up to multiple times to work on the exactness. The pretest power (G power) got is 80%. The dataset utilized is MNIST (Modified National Institute of Standards and Technology). The dataset contains an aggregate of 70000 pictures of various digits from 0 to 9. Train set contains 60000 pictures, and the test set contains 10000 pictures [11]. KNN [7] is a calculation which is utilized for relapse and arrangement purposes. It is a sluggish AI calculation. Euclidean distance is utilized by the KNN calculation to discover the distance between reference focuses to that of the testing focuses which will be useful to find K-esteem.

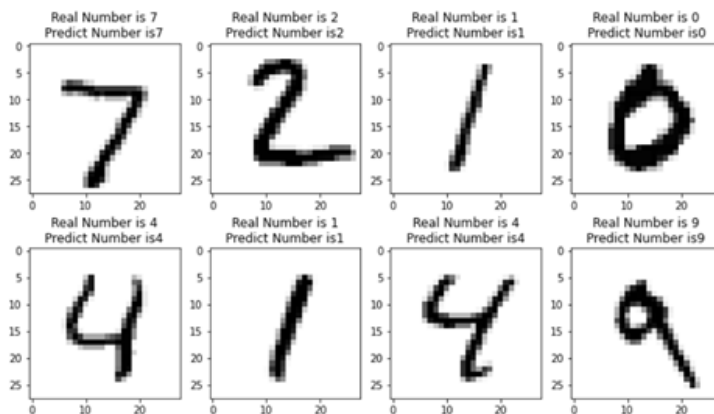
NB calculation [5] is utilized for order purposes. The calculation doesn't have any familiarity with a specific element in a class that is inconsequential to the element of another class. The calculation can be performed on enormous datasets and can be effectively constructed. It can anticipate quick based on the test set. The characterization is free among the suppositions of indicators. Both the algorithms KNN and NB are evaluated in google colab. The hardware configurations were i5 processor, 8 GB Ram, Windows 10 Operating system.

**3. Statistical Analysis**

The examination of the calculations was done in a factual programming device IBM SPSS adaptation 21 [11]. The autonomous factors present in the dataset is size of the picture and ward factors are picture pixels and picture names. The calculations were assessed with the example size of 25 and were iterated for 15 ages to get the expanded exactness. With the got values for every age the Independent Sample t-Test was performed and the diagrams are plotted to examine the exactness of the calculation. In the NB calculation, it is seen that for each cycle the exactness of the digit ID expanded.

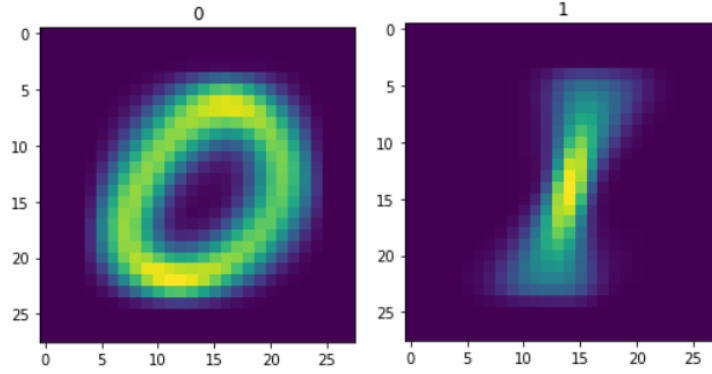
**4. Results and Discussion**

The result of perceiving transcribed characters by utilizing KNN calculation is displayed in Fig. 1, the correlation of the genuine number to the normal number. Barely any digits are referenced in Fig. 1.

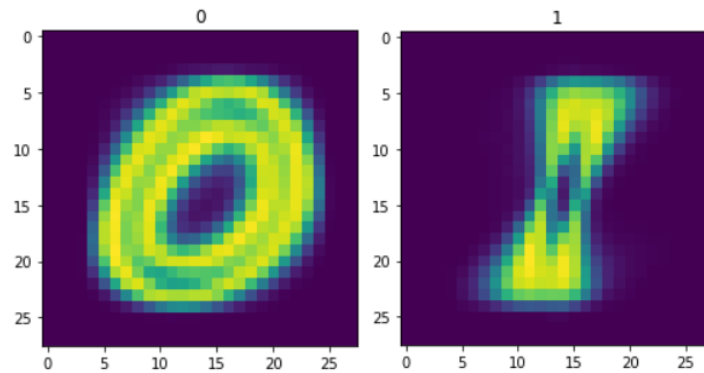


**FIGURE 1.** The outcome of the KNN algorithm in conducted experiments.

As the blunder rate diminishes the precision in foreseeing the number increments. As the quantity of ages expands the precision increases.

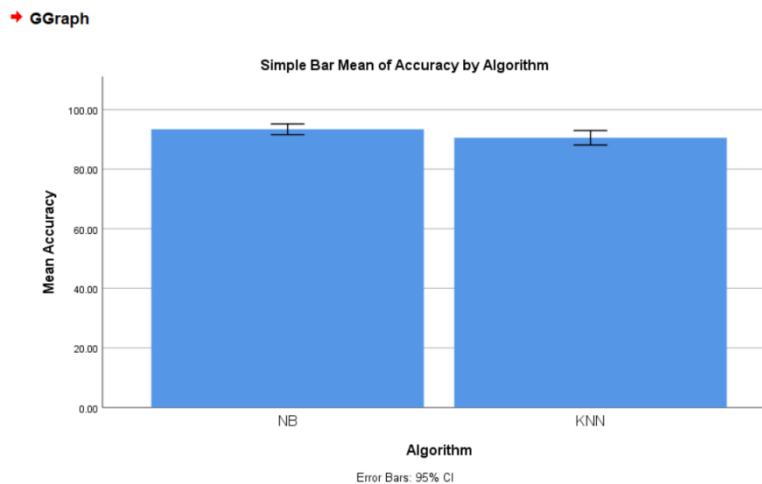


**FIGURE 2.** The mean of every digit from 0-9 using the NB algorithm in the experiment.



**FIGURE 3.** The covariance of every digit from 0-9 using the NB algorithm and its illustration.

The mean and covariance of the digits utilizing the NB calculation is displayed in Fig. 2 and Fig. 3. As the mean and covariance of the digits help the calculation in perceiving the characters even more effectively. Figure 4 addresses the examination of the calculation with individual of the correctnesses including the mistake rate. The diagram shows the NB calculation has an altogether more exactness rate than the KNN calculation.



**FIGURE 4.** The comparison between NB and KNN algorithms including error rate. Mean accuracy of NB algorithm is higher than KNN algorithm. X Axis: KNN vs SVM algorithms. Y Axis is a mean accuracy comparison of KNN and SVM algorithms  $\pm 1$  SD.

Table 1 addresses the mean happened by the NB calculation is 93.3913 and by the KNN calculation is 90.5418 for the 15 age values. NB calculation accomplished more exactness than the KNN calculation and the mistake pace of NB calculation is 0.84478 and the blunder pace of KNN calculation is 1.13782. The blunder pace of NB is lesser than the KNN calculation for perceiving the transcribed characters. From Table 1, it is presumed that the importance esteem got by the NB calculation is 0.165. In this way, the NB calculation seems to have more powerful outcomes than the KNN calculation. Certainty time period distinction of both the calculations is allotted as 95%. Henceforth this transcribed person acknowledgment calculation will be gone about as a clever methodology, and it is authored as Novel Character Recognition (NCR).

**TABLE 1.** The mean accuracy and standard mean rate of NB and KNN algorithms is listed.

	Algorithm	N (Number of Epochs)	Mean	Standard Deviation	Standard Mean Error
Accuracy	NB	15	93.3913	3.27183	0.84478
Accuracy	KNN	15	90.5418	4.40676	1.13782

The examination was done on the NB and the KNN calculations to foresee transcribed person acknowledgment with exactness execution measures. The NB calculation gives off an impression of being superior to the KNN calculation. As the critical worth is more than 0.05, the NB calculation seems to create more compelling outcomes than KNN calculation. The NB calculation with the example size of 25 and the KNN calculation with the example size of 25 were iterated multiple times to create better exactness and accuracy. The expectation of the NB calculation seems, by all accounts, to be higher than the forecast of the KNN calculation. The standard mean was determined by the calculations for 15 emphases and furthermore Standard deviation and Standard mean mistake for the calculations has been classified. The elements influencing the manually written person acknowledgment is picture quality, dataset size, obscured pictures in the dataset. The impediments are the NB has higher time intricacy than the KNN calculation, various shapes and letter sets, colossal changeability and uncertainty of strokes, low quality of source archive, cursive penmanship makes detachment and acknowledgment of characters [10]. Hugh change is utilized for revising slanted filtered pages and fragmented characters. The greater dataset invests in some opportunity for the KNN calculation. The NB calculation has a lower mistake rate than the KNN calculation. On the off chance that the mistake rate is diminished generally, the outcomes would be substantially more compelling. Various calculations have different expectation rates.

## 5. Conclusions

In this examination explore, the exactness result of the NB calculation is 97.45% and which has all the earmarks of being superior to KNN calculation, since it has precision of 97.3% for perceiving the transcribed characters. As the NB calculation seems, by all accounts, to be higher precision, it will be doable to broaden work further on the NB calculation than KNN calculation.

## References

1. N. Meghanathan, D. Nagamalai, and N. Chaki, "Advances in computing and information technology," in Proceedings of the second international conference on advances in computing and information technology, (IEEE, 2012), pp. 750–754.
2. U. Tiwari, M. Jain, and S. Mehfuz, Handwritten character recognition—an analysis, in Advances in System Optimization and Control-2019, Springer, Singapore, 207-212 (2019).
3. P. V. Bhagyasree, A. James, and C. Saravanan, "A proposed framework for recognition of handwritten cursive english characters using DAG-CNN," in 1st International Conference on Innovations in Information and Communication Technology-2019, (IEEE, 2019 April), pp. 1-4.
4. K. Peymani, and M. Soryani, "From machine generated to handwritten character recognition; a deep learning approach," in 3rd International Conference on Pattern Recognition and Image Analysis-2017,(IEEE, 2017 April), pp. 243-247.
5. N. Araki, et al., "A statistical approach for handwritten character recognition using bayesian filter," in 3rd International Conference on Innovative Computing Information and Control-2008,(IEEE, 2008).
6. U.R. Babu, Y. Venkateswarlu, and A.K. Chintha, "Handwritten digit recognition using K-nearest neighbour classifier," in World Congress on Computing and Communication Technologies-2014, (IEEE, 2014 February), pp. 60-65.
7. M. Kumar, M.K. Jindal, and R.K. Sharma, "k-nearest neighbor based offline handwritten Gurmukhi character recognition," in International conference on image information processing-2011, (IEEE, 2011 November), pp. 1-4.
8. P. Ghadekar, S. Ingole, and D.Sonone, "Handwritten digit and letter recognition using hybrid dwt-dct with knn and svm classifier," in Fourth International Conference on Computing Communication Control and Automation-2018, (IEEE, 2018 August), pp. 1-6.
9. M. Rajnoha, R. Burget, and M.K. Dutta, "Offline handwritten text recognition using support vector machines," in 4th International Conference on Signal Processing and Integrated Networks-2017, (IEEE, 2017 February), pp. 132-136.
10. J. Kritika, et al., Ankyloglossia with cleft lip: A rare case report, Journal of Indian Society of Periodontology **19.6**, 690 (2015).
11. M. Ahad, and S. Gheena, Awareness, attitude and knowledge about evidence based dentistry among the dental practitioner in Chennai city, Research Journal of Pharmacy and Technology **9(11)**, 1863-1866 (2016).