DEVELOPMENT AND EVALUATION OF A VISION-BASED SYSTEM FOR GUAVA FRUIT SORTING USING KNN CLASSIFIER

Ashok Kanade[1*], Rajendra Gosavi[2], Arvind Shaligram[3]

[1]*Associate Professor, Department of Electronic Science, P.V.P. College of ASC, Pravaranagar, India, 413713, E-mail:* ashokkanade@gmail.com

[2]*Professor, Department of Electronic-Science, ASC College, Rahuri 413 705, India*
E-mail: rajagosavi967@gmail.com

[3] *Emeritus professor, Department of Electronic-Science, Savitribai Phule Pune University, Pune, India, 413713, E-mail:* adshaligram@gmail.com

ABSTRACT

Guava fruits were categorized into four ripeness levels: Green, Ripe, Overripe, and Spoiled. Classification was achieved through analysis of color features extracted from images. Each ripeness category consisted of 100 fruits. The sorting mechanism employed HSI color space with a K-nearest neighbor (KNN) method. This process used the NI color classification training interface and NI Vision Assistant, utilizing Sum (Manhattan/Taxicab) distance metric to develop a model for a LabVIEW graphical user interface (GUI) to sort fruits on a conveyor system. The HSI+KNN system demonstrated 90% accuracy, compared to 100% accuracy achieved through human grading. The system exhibited 100% repeatability during testing. Observations indicated that sorting speed was primarily limited by the conveyor belt's pace and fruit spacing, rather than the computational time of the vision system, which consistently processed fruit at less than 40 ms per fruit, while mechanical handling required 1000–5000 ms per fruit. Potential factors influencing system performance, such as lighting conditions, camera quality, exposure settings, and camera distance from the fruit, were identified as critical for future optimization. These findings suggest the need for enhanced training datasets, improved imaging control, and integration of multiple methods to achieve superior fruit grading.

Keywords: *Guava fruit sorting, Vision-based system, KNN classifier, HSI color space, Ripeness levels, Machine vision, Color features*

1 Introduction

Quality evaluation is critical in fruit production, as accurate ripeness assessment reduces postharvest losses and supports harvest timing, market strategy, and storage decisions[1]. Among climacteric fruits, guava stands out for its nutritional value as a vitamin C source. Guava has high respiration rates and very short postharvest life, limiting transportation and storage [2]. Conventional guava sorting relies on trained human sensory panels, a method susceptible to to subjectivity, lack of repeatability, and fatigue [3].Previous studies have examined some precise yet harmful physicochemical analysis techniques. [4]. While both e-nose and physicochemical analysis deliver greater accuracy, their lengthy evaluation periods hinder rapid online assessment, rendering them appropriate solely for offline random sampling. While machine vision systems provide potential solutions for fruit sorting, a fast, low-cost, and objective online system specifically tailored to guava's rapid postharvest deterioration remains a critical need [4], [5].

This study investigates the feasibility of a fast, low-cost, online machine-vision system for guava fruit grading, hypothesizing that it can achieve comparable accuracy to human grading while offering superior objectivity and repeatability compared to existing manual and slower analytical methods. The performance of the developed systems was analyzed experimentally for pre-packaging classification of Guava fruits

2. MATERIALS AND METHODS

2.1 Image acquisition and illumination hardware

The iBall 's 12-megapixel webcam is used for image acquisition. This camera has six white LED for illumination. It incorporates six white LEDs for illumination, with brightness controllable via a potentiometer on the interface cable. A LUX meter is used to adjust this brightness. The LEDs originally had a color temperature near that of white light. The spectral characteristics of illuminating source is characterized using stellar Net miniature spectrometer systems and software. The spectral characteristics is shown in figure 1.

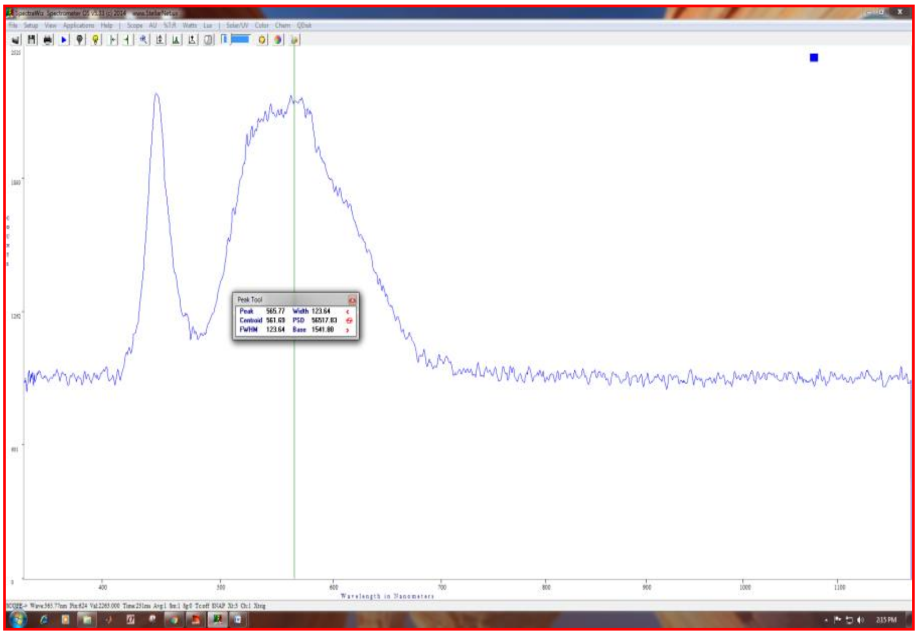

Figure 3. 1 Spectral characteristics of illumination source

This plot indicates two peaks: the lower peak has a peak wavelength of 445.88 nm, and the upper peak has a peak wavelength of 565.77 nm. The CCT in the range of 5000 ^{o}K to 8000^{o}K. This illumination system provides uniform light intensity across the fruit sample plane, as shown in Figure 2.

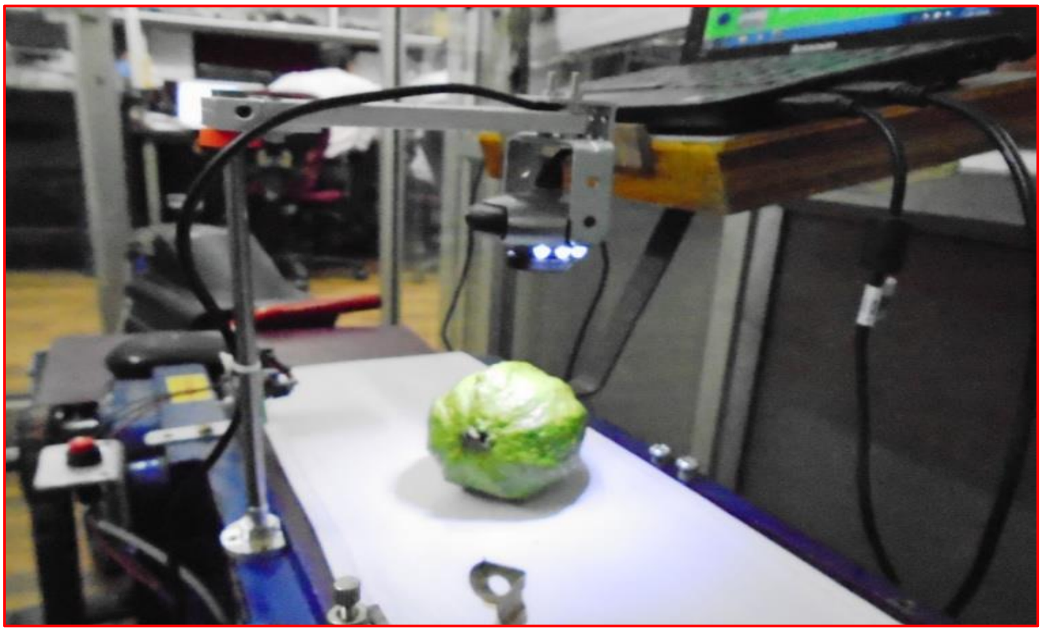

Figure 2 Fruit under webcam with integrated light illumination source

A digital web camera, the iBall i12, was positioned vertically 10 cm from the sample. The angle between the camera lens axis and the lighting sources was approximately 45 °. The experimentation was conducted within a closed room with dark-painted internal walls to minimize light reflection.

2.2 K-Nearest Neighbor (KNN) Classifier

The nearest-neighbor classifier is a non-parametric classification method that does not assume any particular form for the conditional density of a class. This algorithm assigns a pattern label based on the most frequently represented label among the K-nearest samples in the training set, where K denotes the number of nearest neighbors. The entire training dataset is utilized to assign a class to a test pattern, necessitating the measurement of distances between the test pattern and each training pattern[6], [7]. This approach requires substantial computer memory, as it must retain information for every sample in the training set. Various metrics are typically employed to calculate the similarity between patterns. In this study, the Sum metric, also known as the Manhattan or Taxicab metric, is used to measure the distance (similarity) between the test pattern and sample patterns in the training set. In the 1-NN classifier, the test pattern is assigned to the class containing the pattern (training) closest to the measured pattern (testing). Additionally, 2-NN and 3-NN classifiers were evaluated; in these classifiers, the pattern is assigned to the class with the majority of patterns represented within K-NN. When each of the nearest-neighbor patterns (training) belongs to a different class in 2-NN and 3-NN classifiers, the test pattern is assigned to the class with the closest member to the test pattern, i.e., the 1-NN classification procedure is applied when each nearest-neighbor pattern is from a different quality class in 2-NN and 3-NN classifiers[8], [9]. The taxicab distance d1 between two vectors p and q in an n-dimensional real vector space with a fixed Cartesian coordinate system is the sum of the lengths of the projections of the line segment between the points onto the coordinate axes. More formally expressed as,

$$d_1(p,q) = \|p - q\|_1 = \sum_{i=1}^{n} |p_i - q_i| \qquad (1)$$

Where p and q are vectors

$p = (p_1, p_2, \ldots, p_n)$ and $q = (q_1, q_2, \ldots, q_n)$

2.3 Image preprocessing and training ANN

In the present study, a neural network was trained using a novel NI color classification training interface. The color classifier utilized color features to identify samples. The NI color classification training interface is based on IMAQ Train Nearest Neighbor VI as shown in fig. 3

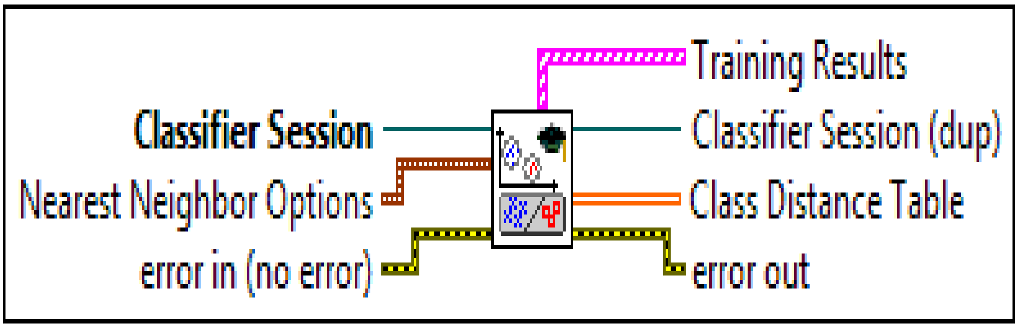

Figure 3. IMAQ Train nearest Neighbor VI

This VI sets the classifier session for the Nearest Neighbor Classifier engine and configures its parameters. Several color classification algorithms are available in the NI Color Classifier, including Nearest Neighbor, K Nearest Neighbor, and Minimum Mean Distance; however, Nearest Neighbor is preferred for this application based on our experimental experience. Nearest Neighbor-based classification represents the most direct approach for classification[10]. In nearest neighbor classification, the distance between an unknown color sample and other classes is defined as the distance to the closest samples within those classes.

The classification score reflects the certainty with which a sample is assigned to one class rather than another. The identification score measures the similarity between a sample and samples within the assigned class. The closest sample image represents the learned sample that most closely matches the sample within the region of interest. The proper option from the metric list was selected to configure how the classification engine calculates the distance between samples. Sum—Metric used in the present applications. Figure 4 illustrates the complete steps involved in training the KNN using the NI color classification training interface.

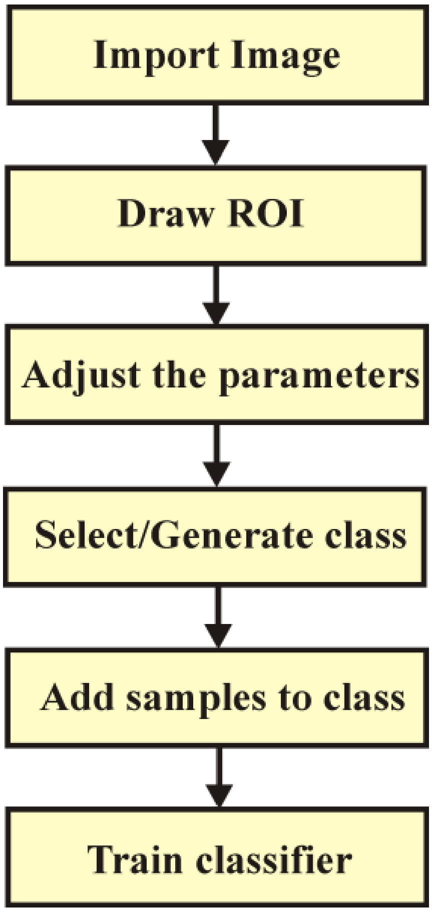

Figure 4 Flowchart for use of NI color classification training interface

The four guava fruit classes–green, ripe, overripe, and spoiled–were created using the NI color classification training interface, as illustrated in Figure 5.

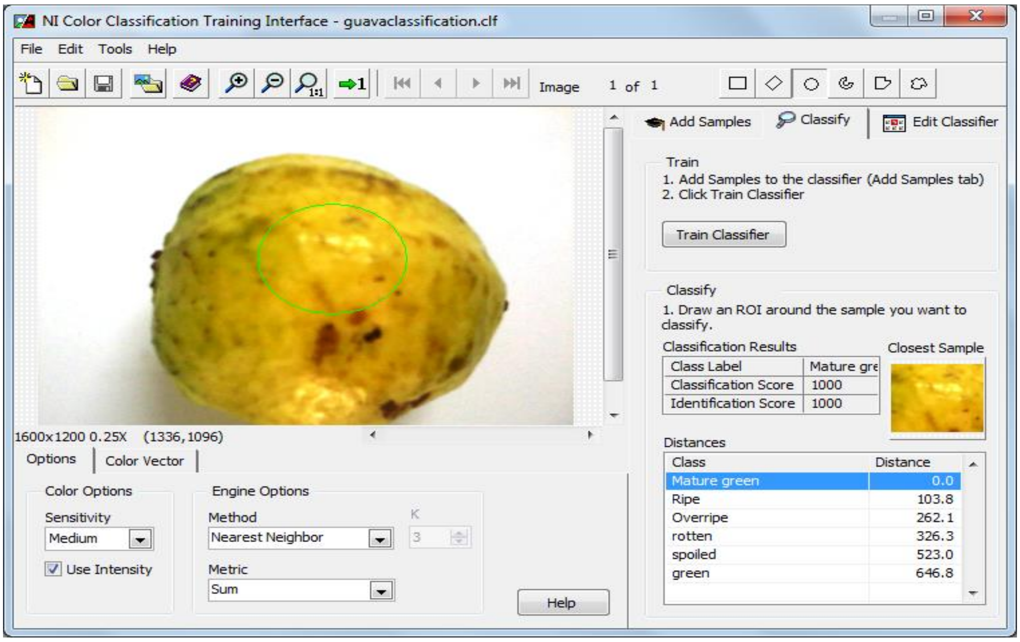

Figure 5 NI color classification training interface

Guava fruits sorted by a trained human classifier were used for training, and the resulting trained color classifier file was generated. This file was integrated into the graphical user interface (GUI) of the developed software. Furthermore, the classifier file underwent verification within the NI Vision Assistant using a script designed for guava fruit classification based on color, as illustrated in Figure 6.

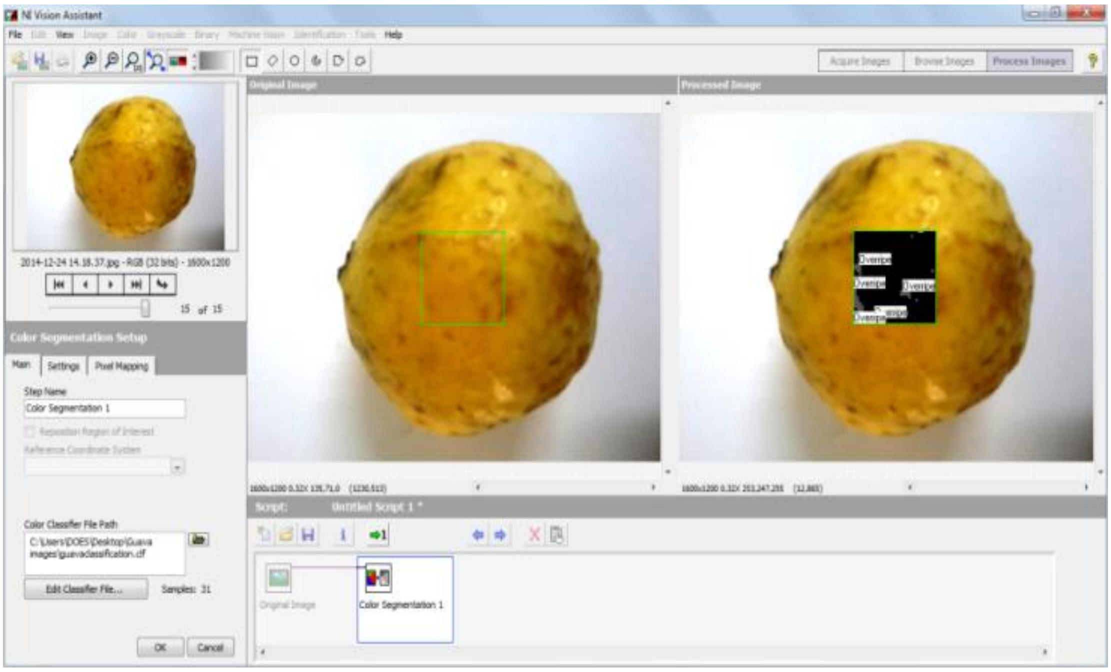

Figure 6: Testing of the trained classifier within the NI Vision assistant.

2.4. Experimental design

The guava fruit was selected for this study. Guava fruits were classified into four groups: Green, Ripe, Overripe, and Spoiled. One hundred fruit samples from each ripening class were utilized for experimentation. These fruit samples were divided into training and testing sets, with 50 % of the fruit images from each group were used to train the PARC system and the remaining 50% fruits serve as the testing set. Images of the training set samples were collected at various times of day and from different angles within each category. These variations increased data set variability and represented a more realistic scenario. The Hierarchical Spatial Index (HSI) color space and k-nearest neighbor classifier-based pattern recognition modules were trained and employed in this study. Each fruit was passed through a conveyer belt at different times of day, and the classification results were recorded.

3. Results And Discussion

The machine vision-based fruit sorting system developed is employed for the classification of guava fruits into four distinct categories at the pre-packaging stage. The GUI is shown in figure 7. The software employs a trained classifier file, previously prepared using the National Instruments (NI) color classification training interface as described previously. This intelligent grading software has been practically applied to classify guava fruits during the pre-packaging phase.

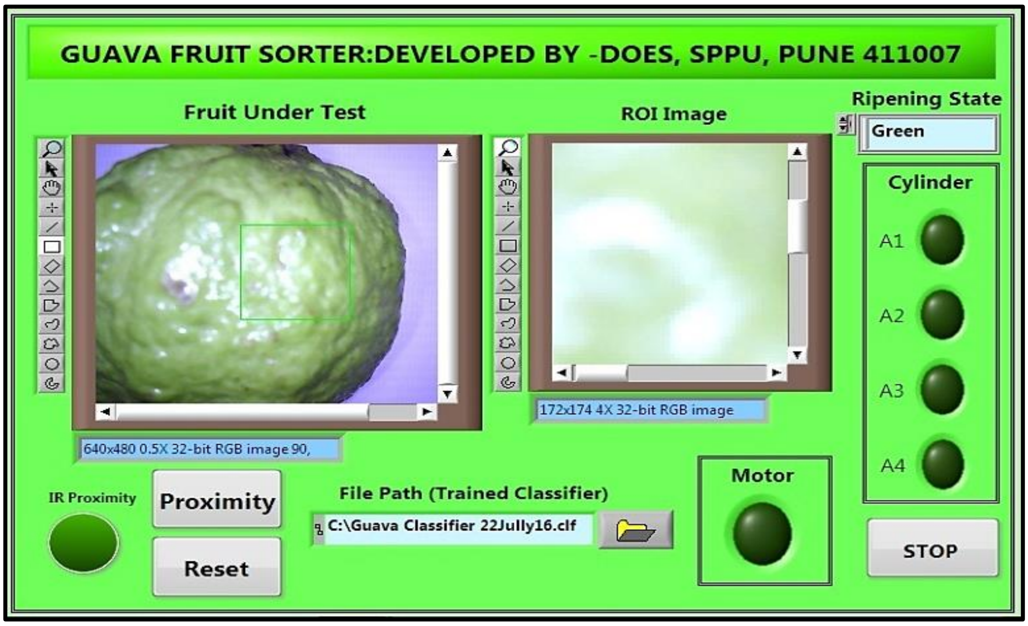

Figure 7 K nearest neighbour based classifier window

The classification accuracy of the developed system for the four categories of guava fruit is presented in Table 1. The classifier demonstrates an average efficiency of approximately 90%, assuming human grading, serving as the reference standard, is 100% accurate. It is noteworthy that the 10% discrepancy arises from the subjective judgment of human graders in evaluating the ripening stage of the fruit during manual grading–a factor inherently unavoidable. The system's repeatability has been confirmed at 100% through extensive experimental validation.

Table 1 Performance analysis of k- nearest neighbour-based PARC

Ripening state	**No. of samples**		**Classification**		**performance %**
	Training	**Testing**	**Expert**	**System**	
Green	50	50	50	50	100
Ripe	50	50	50	40	80
Overripe	50	50	50	44	88
Spoiled	50	50	50	46	92

4. Conclusions

The developed fruit sorting system demonstrates efficacy in categorizing guava fruits at various ripening stages through analysis of their color features. Empirical evaluations indicate that the Nearest Neighbor Classifier, utilizing the Sum Metric, yields favorable outcomes for this task. System performance improves with an increase in the number of training samples. Assessing the system's accuracy percentage reveals that it effectively translates human visual assessment of guava fruit classification into a machine vision process. The system consistently identifies the correct category of guava fruit when tested with the complete set of fruits used in both the training and classification phases. This proposed system holds significant potential in the domain of machine vision-based fruit classification and grading. Applying machine vision techniques, including image processing and various Photoperiod Area Ratio (PAR) algorithms, for the automatic sorting of fruits is rapid, cost-effective, and intelligent. The sorting system's speed is constrained by the conveyor belt's pace and fruit spacing, rather than the computerized vision system's response time. Given that the algorithms are software-based, the system can be adapted to inspect other fruits and agricultural products. The study investigated variations in grading performance due to factors such as ambient lighting, camera resolution, exposure time, and camera distance from the fruit. The findings suggest that the machine vision-based system's performance closely aligns with that of human experts, who evaluate fruits based on skin color, size, surface defects, shape, firmness, weight, and smell.

REFERENCES

[1] B. Li, J. Lecourt, and G. Bishop, "Advances in Non-Destructive Early Assessment of Fruit Ripeness towards Defining Optimal Time of Harvest and Yield Prediction—A Review," *Plants*, vol. 7, no. 1, p. 3, Jan. 2018, doi: 10.3390/plants7010003.

[2] S. Bangayya and A. Kanade, "RGB color space based machine vision system for Classification of guava fruits," *Journal of the Maharaja Sayajirao University of Baroda*, vol. 55, no. 2, pp. 496–499, 2021, doi: 10.5281/zenodo.18287380.

[3] L. Sipos, Á. Nyitrai, G. Hitka, L. F. Friedrich, and Z. Kókai, "Sensory Panel Performance Evaluation—Comprehensive Review of Practical Approaches," *Applied Sciences*, vol. 11, no. 24, p. 11977, Dec. 2021, doi: 10.3390/app112411977.

[4] Y. Shi *et al.*, "A Review on Meat Quality Evaluation Methods Based on Non-Destructive Computer Vision and Artificial Intelligence Technologies," *Food Sci Anim Resour*, vol. 41, no. 4, pp. 563–588, July 2021, doi: 10.5851/kosfa.2021.e25.

[5] X. Huang *et al.*, "Rapid and nondestructive detection of freshness quality of postharvest spinaches based on machine vision and electronic nose," *Journal of Food Safety*, vol. 39, no. 6, Oct. 2019, doi: 10.1111/jfs.12708.

[6] M. Khan, Q. Ding, and W. Perrizo, "k-nearest Neighbor Classification on Spatial Data Streams Using P-trees," Springer Berlin Heidelberg, 2002, pp. 517–528. doi: 10.1007/3-540-47887-6_51.

[7] J. Huang, Y. Wei, J. Yi, and M. Liu, "An Improved kNN Based on Class Contribution and Feature Weighting," Institute Of Electrical Electronics Engineers, Feb. 2018, pp. 313–316. doi: 10.1109/icmtma.2018.00083.

[8] N. Hidayati and A. Hermawan, "K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation," *JEATech*, vol. 2, no. 2, Aug. 2021, doi: 10.21831/jeatech.v2i2.42777.

[9] Y. Song, J. Huang, D. Zhou, H. Zha, and C. L. Giles, "IKNN: Informative K-Nearest Neighbor Pattern Classification," Springer Berlin Heidelberg, 2007, pp. 248–264. doi: 10.1007/978-3-540-74976-9_25.

[10] C. Feng, B. Zhao, X. Zhou, X. Ding, and Z. Shan, "An Enhanced Quantum K-Nearest Neighbor Classification Algorithm Based on Polar Distance," *Entropy*, vol. 25, no. 1, p. 127, Jan. 2023, doi: 10.3390/e25010127.