International Journal of Membrane Science and Technology, 2023, Vol. 10, No. 5, pp 895-903

Improving Kinnow Fruit Classification with Feature Weighting using Modified Levy Flight Distribution-Inspired Sunflower Optimization Technique

Sukhpreet Singh [1], Kamal Malik [2]

[1] *Department of Computer Science and Engineering, CT University, Ludhiana, Punjab, India*
sukhpreet.manshahia@gmail.com

[2] *Department of Computer Science and Engineering, CT University, Ludhiana, Punjab, India*
kamal.malik91@gmail.com

Abstract: In this paper, Feature weighting is used to create an intelligent and effective classification method for Kinnow fruits. Feature weighting approach is used because it improves classification performance more than feature selection methods. The modified sunflower optimization algorithm (SFO) is proposed to search the optimal feature weights and parametric values of k-Nearest Neighbour (kNN).The levy flight distribution operator has been utilised to enhance the convergence speed of the sunflower optimization algorithm by improving the local and global search ability of the optimization algorithm. Also, the algorithmic parameter of the SFO algorithm has been adaptively selected using the linear time varying adaption method. In addition, tanh normalization technique is used for the data pre-processing to reduce the influence of outliers and dominating features before the feature weighting method. The findings suggest that the proposed wrapper based approach feature weighting technique is more capable of achieving higher accuracy than the existing strategies.

Keywords: Feature weighting, Intelligent classification, Optimization technique, Wrapper approach.

1. INTRODUCTION

The Kinnow fruit holds a paramount position in the citrus agriculture of India, being both the most widely cultivated and consumed citrus variety. It finds cultivation in approximately half of the citrus-producing regions within the country. Kinnow fruit stands out due to its exceptional production, remarkable juice content, and outstanding fruit quality, surpassing other citrus crops in these attributes. The primary cultivation area for the Kinnow citrus variety is the Punjab region. With its characteristic orange hue and spherical shape, it typically ripens during the months of January and February. Owing to its substantial vitamin C and calcium content, Kinnow fruit exerts a profound economic influence [1].

The post-harvest phase, which encompasses the intricate procedures of fruit handling and packaging, is inherently interwoven with the precise discipline of fruit grading. Within the specific domain of Kinnow fruit cultivation, the process of grading assumes a pivotal role, imposing a critical responsibility on the agricultural community. Traditional manual grading methods, while prevalent, exhibit susceptibility to human errors, resulting in heightened time consumption and increased economic expenditure.

In the modern era, the adoption of computer vision techniques for automated inspection of fruits has gained prominence as a means of categorizing fruits into distinct groups, offering the dual benefits of reducing labour costs and enhancing product quality [2]. The automated categorization of fruits carries significant importance, as it has the potential to bolster growers' revenue. Moreover, the application of classification techniques in fruit post-harvesting holds promise for substantial improvements. Several researchers have contributed to this field, proposing diverse methods for defect categorization and quality assessment of Kinnow fruits. Singh and Nill, for instance, introduced a method for Kinnow classification based on color-based attributes, utilizing the k-means segmentation technique and an artificial neural network classifier [3]. Hadimani and Mittal, on the other hand, presented a computer vision algorithm that leverages color and physiochemical parameters to determine the ripeness of Kinnow fruits [4]. Yadav et al. conducted a comprehensive examination of the physical characteristics of Kinnow fruits for the purpose of grading [5]. Furthermore, Hadimani and Garg introduced a method for categorizing various flaws in Kinnow fruits, employing multiple criteria for classification [6]. Akhter et al. proposed a classification methodology for categorizing citrus fruits, including Kinnow, by utilizing color and shape data, and incorporating principal component analysis for feature extraction [7]. Additionally, Singh and Malik contributed to the field by suggesting an approach using feature selection for enhancing the classification of Kinnow fruits, employing the hybrid filter method [8].

Existing classification methods for Kinnow performance warrant enhancement, primarily because of their limited consideration of the impact of redundant and irrelevant features on the classification process. An additional challenge is posed by the presence of outliers and the prominence of dominating features, which collectively undermine the accuracy of the classification task. Furthermore, the judicious selection of optimal parameters for the classifier assumes paramount significance to attain heightened performance standards.

Motivated by the intricate challenges associated with Kinnow fruit classification, our research endeavors aim to contribute a comprehensive approach that systematically addresses the identified issues. Our methodology is carefully designed to not only recognize and understand the complexities of the classification process but also to provide effective solutions.

Central to our approach is the strategic selection of optimal feature subsets including classifier parameter values. This step is critical as it involves identifying the most relevant features and tuning the classifier parameters to ensure optimal performance. The goal is to create a refined and effective model to accurately categorize Kinnow fruits based on their diverse characteristics.

A key focus of our methodology is addressing the outlier’s presence and dominating characteristics influence. The accuracy of the classification process can be significantly impacted by these factors. To counteract these challenges, we incorporate a normalization method. This method plays a crucial role in standardizing the data, mitigating the outlier’s influence, and ensuring a more balanced representation of features. In this manner, our goal is to strengthen the durability and dependability of the classification system.

In essence, our approach is a holistic strategy that combines meticulous feature selection, parameter optimization, and normalization techniques to boost the overall effectiveness of Kinnow fruit classification. Through these efforts, we aspire to contribute to the evolution of accurate and reliable fruit grading systems.

2. SUNFLOWER OPTIMIZATION ALGORITHM

Gomes et al. have previously advocated the utilization of the Sunflower Optimization (SFO) algorithm for the deterioration detection in laminated composite plates [9]. This algorithm functions as a part of a class of techniques inspired by biological processes and is specifically rooted in the behaviour and orientation of sunflowers as they track the sun's position. The SFO algorithm lures its inspiration from the biological phenomenon of sunflowers' movements in response to the sun's motion, a natural mechanism employed by sunflowers to optimize their exposure to solar energy. This innovative approach leverages principles from the natural world to address real-world engineering challenges.

Sunflowers exhibit a fascinating phenomenon known as heliotropism, a term derived from their natural inclination to move in the sun’s direction [10]. Throughout the day’s duration, sunflowers display this distinctive behaviour by initially tilting their heads towards the east before the sun rises. As daylight progresses, they meticulously track the sun's movement across the sky until it eventually sets in the west. During the night, sunflowers reorient themselves, facing east once more, in preparation for their sun-following ritual at the break of dawn. The Sunflower Optimization (SFO) algorithm is ingeniously inspired by these sunflower behaviours and their efficient utilization of solar radiation. It draws from the principles of natural evolution within sunflowers, resulting in a unique and bio-inspired approach for problem-solving.

Furthermore, the Sunflower Optimization (SFO) algorithm incorporates a fundamental principle known as the inverse square law of radiation. This law posits that the strength or quantity of radiation is inversely proportional to the square of the distance from the source. In simple terms, the square of the distance between objects governs the amount of radiation they receive or emit. This concept is integral to the SFO algorithm, as it encapsulates the way sunflowers optimize their exposure to sunlight based on their distance from the sun, and it serves as a key mechanism within the algorithm for optimizing solutions to complex problems.

The amount of heat H that each plant absorbs is represented by the equation below:

$$H_l = \frac{Po}{4\pi r_l^2} \quad (1)$$

Po is the source of power, and $r_l$is the distance between the current source and plant l. Sun direction for sunflowers is as follows:

$$\overrightarrow{Sn_l} = \frac{P^*}{\|P^* - P_l\|} \quad l = 1,2,3, \dots, n \quad (2)$$

where the best value from each iteration and throughout the whole population, respectively, is denoted by P_l and P^*. The sunflowers' stride in the direction sn_l is determined by:

$$dis_l = \gamma * R_l(\|P_l - P_{l-1}\|) * \|P_l + P_{l-1}\| \quad (3)$$

where γ is the inertial displacement and the likelihood of the pollination is indicated by $\|P_l - P_{l-1}\|$.

The maximum restriction for the step for every individual is established below to avoid bouncing zones that considered as global minimums:

$$dis_{max} = (\|P_{max} - P_{min}\|) * 2 * Pop_n \quad (4)$$

Pop_n is the total number of plants, P_{max} and P_{min} are upper and lower limits, respectively.

$$\overrightarrow{P_{i+1}} = \overrightarrow{P_i} + dis_l * \overrightarrow{Sn_l} \quad (5)$$

3. PROPOSED MODIFIED LEVY FLIGHT DISTRIBUTION-BASED SUNFLOWER OPTIMIZATION ALGORITHM

In this algorithm, an innovative method known as the Modified Levy Flight distribution-based Sunflower Optimization (MLFSFO) algorithm is proposed. This pioneering approach stems from the traditional Sunflower Optimization (SFO) algorithm, integrating the Levy Flight distribution operator to expedite convergence speed. The incorporation of the Levy Flight operator is strategically designed to fortify the algorithm's performance, fostering improvements in both local and global search capabilities. Through this enhancement, our objective is to create a more vigorous optimization framework that can efficiently navigate complex solution spaces, thereby advancing the algorithm's overall effectiveness.

The Levy Flight operator plays a pivotal role in enabling plants, akin to sunflowers, to explore their surroundings more effectively. It facilitates rapid movements while occasionally allowing for substantial leaps, which proves advantageous for the algorithm's capacity to conduct comprehensive searches, encompassing both global and local aspects of the optimization problem at hand.

The following equation has been assessed to produce random numbers from the Levy distribution[11]:

$$L = \frac{\mu}{|v|^{1/\beta}} \tag{6}$$

where μ and v follow the Gaussian distribution having following formulation:

$$\mu \sim \left(0, \sigma_\mu^2\right), \nu \sim (0, \sigma_v^2) \tag{7}$$

where $\sigma_\mu^2 = \left[\frac{\Gamma(1+\beta) * \sin(\pi\beta/2)}{\Gamma\left[\frac{1+\beta}{2}\right]\beta * 2^{\beta - 1/2}}\right]^{1/\beta}$, $\beta = 1.5, \sigma_v = 1$, and Γ is the standard gamma function.

Moreover, in the framework of the MLFSFO algorithm, we delve into the critical aspects of pollination rate and mortality rate through a meticulous adaptive selection process. These algorithmic parameters stand as indispensable elements exerting a substantial impact on the overall efficacy of the MLFSFO algorithm. Their significance lies in their pivotal role in orchestrating and guiding the search process inherent to meta-heuristic optimization algorithms. By dynamically adjusting these parameters, our approach intends to enhance the adaptability and responsiveness of the MLFSFO algorithm, optimizing its performance in navigating complex optimization landscapes.

The optimal configuration of algorithmic parameters plays a pivotal role in shaping the efficacy of the optimization algorithm. In light of this, our methodology places significant emphasis on a systematic evaluation process. We meticulously assess a spectrum of potential parameter combinations, adhering to predefined bounds. This rigorous evaluation aims to discern the most advantageous values for the Sunflower Optimization (SFO) algorithm. Through this systematic exploration, we seek to pinpoint the parameter settings that contribute optimally to the algorithm's overall effectiveness, ensuring a finely tuned and high-performing optimization process.

It is significant to observe that the selection of parameter values can be problem-specific, and the parameter values which are fixed may not be universally suitable for all scenarios. There's a risk of premature convergence and becoming trapped in local optima. To address these challenges, this paper presents an enhanced iteration of the SFO algorithm, offering a solution to the issue of algorithmic parameter selection. We employ the linear time-varying adaptation strategy [12] as an adaptive approach to dynamically determine the optimal values of these parameters, thus enhancing the adaptability and problem-solving capability of the MLFSFO algorithm.

The following is the formulation:

$$val(cr_t) = \frac{t_{max} - cr_t}{t_{max}}(val_{max} - val_{min}) + val_{min} \tag{8}$$

The following algorithm explains how the SFO algorithm functions.

Algorithm: SFO algorithm

Step 1: Randomly initialize the n population of sunflowers.

Step 2: Find the best solution (sun, s*) in the starting population. Place every plant so that it faces the sun.

Step 3: Calculate the mortality rate using equation 8. Determine the direction of each plant's orientation vector. Reduce by a factor of m (mortality rate) the number of plants located in direct sunlight using levy distribution operator.

Step 4: Calculate the pollination rate using equation 8. Determine the appropriate next action for each plant. The best P (pollination rate) plants will pollinate in a circumferential pattern around the sun and levy distribution operator.

Step 5: Analyze the newly introduced solutions. If a new solution emerges as the best in the whole population, the sun should be updated.

Step 6: Repeat steps 3 to 5 until current iteration value is less that the maximum iteration.

4. PROPOSED APPROACH

In the framework of our proposed methodology, we strategically incorporate the tanh data normalization technique to effectively alleviate the influence of outliers and dominant features within the dataset. This normalization procedure plays a pivotal role as an essential pre-processing step, ensuring the dataset's resilience against the impact of irregularities and feature dominance. Subsequently, we introduce a tailored modification to the Sunflower Optimization algorithm. This modification is carefully designed to address both feature weighting and parameter optimization intricacies specifically tailored for the k-Nearest Neighbor (kNN) classifier.This innovative algorithmic enhancement stands out as a significant contribution to the overall robustness and efficiency of the classification process. This comprehensive strategy emphasizes the importance of meticulous preprocessing and algorithmic refinement to fortify the model against challenges such as outliers, dominant features, and classification parameter selection, thereby advancing the overall effectiveness of the classification system.

The figure 1 illustrates the block diagram of the proposed approach.

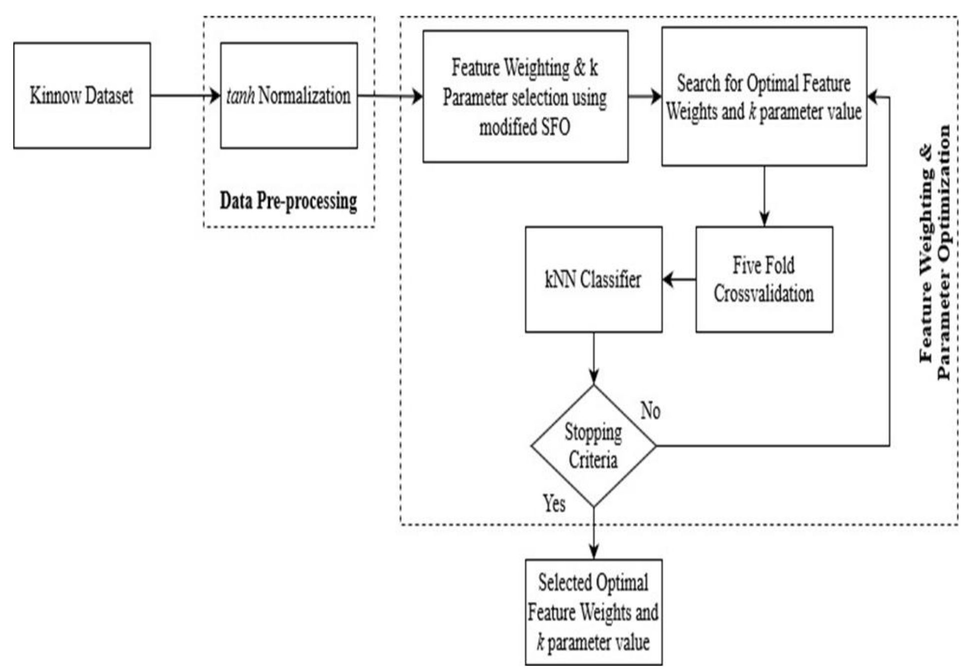

Figure 1: Block diagram of the proposed wrapper based-approach

The details of the proposed wrapper based-approach for classification are described in the following subsections:

4.1 Kinnow Dataset

In this proposed approach, a comprehensive dataset containing a total of 1,200 high-resolution Kinnow fruit images has been carefully assembled. These images represent five distinct categories or classes of Kinnow fruit, each characterized by unique visual attributes and characteristics [8]. The dataset serves as the framework for our research, enabling us to explore and evaluate various aspects of Kinnow fruit classification. There are five kinnow grade categories, namely Grade-A, Grade-B, Grade-C, Grade-D, and Defected, as illustrated in figure 2.

The feature extraction process from these images is a vital step in our analysis. These extracted features encompass a wide array of characteristics, including texture, color, shape, and size. Texture features provide insights into the surface properties of the fruit, allowing us to distinguish different textures that may be indicative of varying ripeness stages or quality. Color features capture the color spectrum of the fruit, helping to discriminate between different hues and shades that are related with specific fruit conditions. Shape features emphasis on the geometric attributes of the fruit, such as its contour and overall form, enabling us to identify distinctive shape patterns within the dataset. Size features quantify the physical dimensions of the fruit, including measurements such as diameter and volume, which can be valuable for differentiation.

By incorporating these diverse feature sets, we aim to design a robust and comprehensive classification model for Kinnow fruit. This model will leverage the unique characteristics extracted from the dataset to effectively categorize Kinnow fruit into their respective classes, providing insights into the fruit's quality, ripeness, and other relevant attributes.

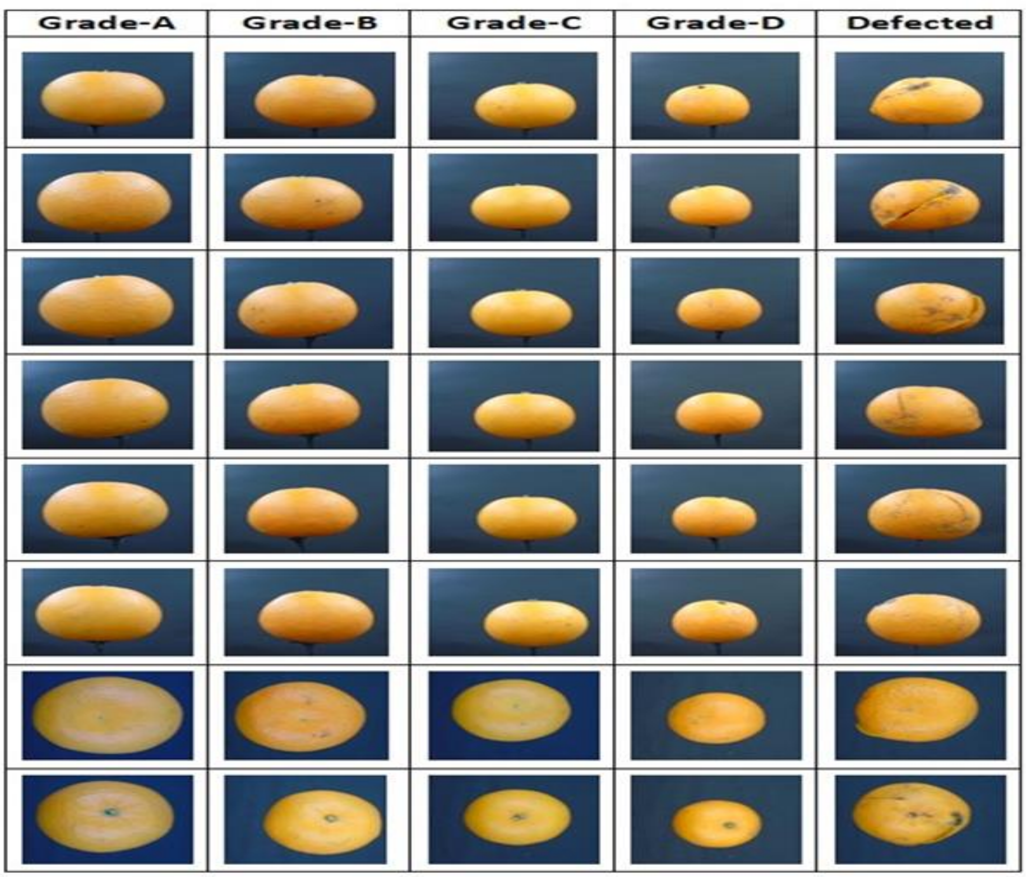

Figure 2: Images of kinnows classified into distinct quality grades.

4.2 Data Pre-processing

Data normalization is an essential pre-processing step employed in this study. Normalization techniques are particularly valuable in addressing outliers and mitigating the impact of feature dominance, especially concerning numerical scales. In our work, we have utilized the hyperbolic tangent (tanh) normalization method, as recommended in the prior research [13], to effectively manage and standardize the dominant feature. By employing this normalization technique, the dataset is standardized to a consistent scale, thereby improving the robustness and accuracy of analyses. This process enhances the performance and reliability of the classification, mitigating the impact of outliers and feature imbalances. Consequently, it contributes to the overall quality of our results and findings.

4.3 Feature Weighting and Parameter Optimization

In our proposed approach, with the aim of enhancing the classification performance of the k-Nearest Neighbors (kNN) classifier, we introduce a feature weighting technique. This technique is rooted in the modified Sunflower Optimization (mSFO) method, which we have designed specifically for this purpose. Additionally, we address the optimal selection of the k parameter in the kNN classifier using our proposed optimization algorithm.

A comprehensive description of our proposed algorithm, including the details of the feature weighting technique and the optimization of the k parameter, is elaborated in Section 3 of this paper. This section provides a thorough exposition of the method's principles and its application within the context of improving the classification accuracy and efficacy of the kNN classifier.

5. RESULTS AND DISCUSSIONS

The modified Sunflower Optimization method is used in this proposed approach for improving the classification of Kinnow fruits. Our experimental endeavors are performed on the MATLAB 2019a platform, utilizing a computer operating under the Windows 10 Pro operating system. The computational infrastructure is powered by an Intel® Xenon® CPU E5-2650 v3 with a clock speed of 2.30 GHz and is equipped with 8 GB of memory.

For the evaluation of our classification performance, we adopt a robust five-fold cross-validation procedure. This method ensures a rigorous and reliable assessment of our proposed approach's effectiveness, enhancing the credibility of our results and findings. The performance of proposed approach has been computed for the thirty reruns, five hundred iterations, and thirty population size.

Figure 3, shows the comparison of the convergence curve of the proposed MLSFO approach and SFO algorithm.

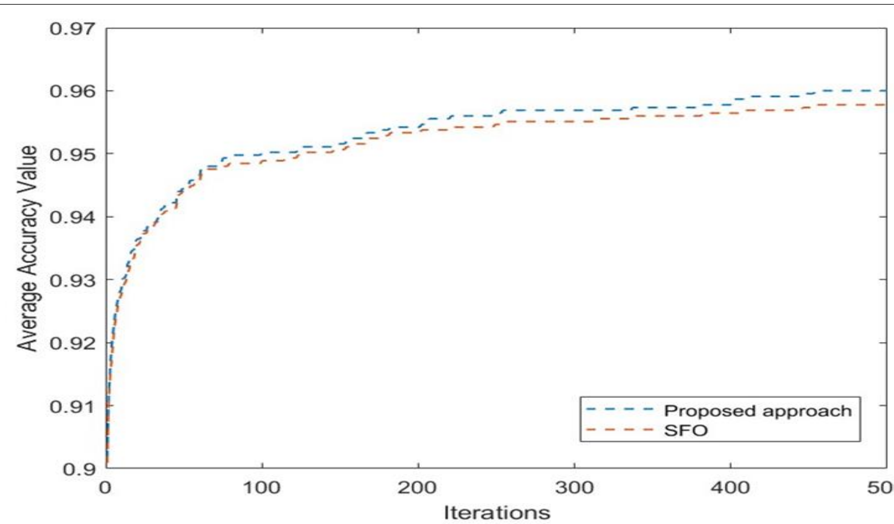

Figure 3: Convergence curve of the SFO and proposed approach

Figure 3 reveals that the proposed approach outperforms the SFO algorithm, demonstrating higher accuracy and faster convergence.

Table 1 presents the mean accuracy and standard deviation for both the SFO and the proposed approach, illustrating the performance of the latter

Table 1. Performance metrics comparison

	Average Accuracy Percentage	Standard Deviation
SFO with *tanh* normalization	95.78 %	0.61
Proposed Approach	96.00 %	0.51

Analysis of Table 1 reveals that the proposed approach outperforms the SFO algorithm in terms of accuracy.

To evaluate the statistical significance of the proposed approach concerning error values, the Friedman's mean rank test was conducted. Remarkably, the proposed approach secures the lowest Friedman's mean rank, scoring below 0.4, indicating its superiority over the SFO algorithm.

6. CONCLUSION

In this work, we present an intelligent approach aimed at enhancing the classification performance of Kinnow fruits. Our proposed methodology combines several key techniques, including tanh normalization and a feature weighting method based on the Modified Levy Flight distribution-based Sunflower Optimization (MLSFO) algorithm. Furthermore, the traditional Sunflower Optimization (SFO) algorithm has been refined through the integration of the Levy flight distribution operator and a linear adaptation method.

The application of the tanh normalization method is instrumental in addressing several critical issues, including the removal of outliers, mitigation of the impact of dominant, redundant, and irrelevant features. These measures collectively lead to a significant improvement in the classification performance of the k-Nearest Neighbors (kNN) classifier. Additionally, our approach optimizes the kNN classifier's parameter settings using the MLSFO algorithm, further enhancing its performance.

In our analysis, we have rigorously evaluated the efficiency of our proposed approach in comparison to the traditional SFO algorithm. The results demonstrate that our intelligent classification approach achieves an impressive accuracy rate of 96.00%.

Looking ahead, the applicability of our proposed approach extends beyond Kinnow fruits, as it holds potential for classifying various other fruits. Future research endeavors may explore the broader application of our intelligent classification methodology to diversify fruit classification efforts.

7. REFERENCES

[1] Samiksha, "Citrus Fruit Cultivation in India-Production Area, Climate, Harvesting and Fruit Handling Citrus Fruit Cultivation in India-Production Area, Climate, Harvesting and Fruit Handling!" 2021.

[2] V. Kavitha and M. R. Devi, "Predicting the Diseases By Graphcut Method for Citrus Fruit," Int. Res. J. Manag. Sci. Technol., vol. 7, no. 12, pp. 465–470, 2016.

[3] H. Singh and N. Gill, "Machine Vision Based Color Grading of Kinnow Mandarin," Int. J. Adv. Res. Comput. Sci. Softw. Eng., vol. 5, no. 5, pp. 1253–1259, 2015.

[4] L. Hadimani and N. Mittal, "Development of a computer vision system to estimate the colour indices of Kinnow mandarins," J. Food Sci. Technol., vol. 56, no. 4, pp. 2305–2311, Apr. 2019, doi: 10.1007/s13197-019-03641-9.

[5] A. Ali, N. K. Yadav, A. Ali, R. Dev, and M. Pandurang, "Evaluation of physical properties of different grades of kinnow mandarin," J. Pharmacogn. Phytochem., vol. 8, no. 1, pp. 1414–1417, 2019.

[6] L. Hadimani and N. M. Garg, "Automatic surface defects classification of Kinnow mandarins using combination of multi-feature fusion techniques," J. Food Process Eng., vol. 44, no. 1, p. e13589, Jan. 2021, doi: 10.1111/jfpe.13589.

[7] N. Akhter, M. Idrees, F. U. Rehman, M. Ilyas, Q. Abbas, and M. Luqman, "Shape and texture based classification of citrus using principal component analysis," Int. J. Agric. Ext., vol. 9, no. 2, pp. 229–238,

Aug. 2021, doi: 10.33687/IJAE.009.02.3525.
[8] S. Singh and K. Malik, “Feature selection and classification improvement of Kinnow using SVM classifier,” Meas. Sensors, vol. 24, no. September, p. 100518, 2022, doi: 10.1016/j.measen.2022.100518.
[9] G. F. Gomes, S. S. da Cunha, A. C. Ancelotti, S. S. da Cunha, and A. C. Ancelotti, “A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates,” Eng. Comput., vol. 35, no. 2, pp. 619–626, 2018, doi: 10.1007/s00366-018-0620-8.
[10] H. S. Atamian, N. M. Creux, E. A. Brown, A. G. Garner, B. K. Blackman, and S. L. Harmer, “Circadian regulation of sunflower heliotropism, floral orientation, and pollinator visits”, Accessed: Jul. 24, 2021. [Online]. Available: http://science.sciencemag.org/
[11] H. Singh, B. Singh, and M. Kaur, “An improved elephant herding optimization for global optimization problems,” Eng. Comput., no. 0123456789, 2021, doi: 10.1007/s00366-021-01471-y.
[12] R. C. Eberhart and Yuhui Shi, “Tracking and optimizing dynamic systems with particle swarms,” in Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546), 2001, vol. 1, pp. 94–100. doi: 10.1109/CEC.2001.934376.
[13] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” Appl. Soft Comput., vol. 97, p. 105524, Dec. 2020, doi:10.1016/j.asoc.2019.105524.

DOI: https://doi.org/10.15379/ijmst.v10i5.3624