International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(RESEARCH ARTICLE)

Check for updates

Machine learning modeling for prediction of sweet potato slices quality based on colour property

Mfrekemfon Godswill Akpan [1,*], Uduak David George [2] and Elijah George Ikrang [1]

[1] *Department of Agricultural and Food Engineering, University of Uyo, Uyo, Nigeria.*
[2] *Department of Computer science, Faculty of Computing University of Uyo, Nigeria.*

International Journal of Science and Research Archive, 2026, 19(03), 1012-1020

Publication history: Received on 15 May 2026; revised on 24 June 2026; accepted on 26 June 2026

Article DOI: https://doi.org/10.30574/ijsra.2026.19.3.1395

Abstract

Processing has effects on some key quality parameters either positively or negatively, there is also need for process monitoring to determine the state of processing. This study was set out to model the colour changes during the drying of Sweet Potato Slices as one of the key parameters of quality. The drying was carried out using two drying methods- hot air drying, HA and Vacuum drying VD. The sweet potato was dried at 3 mm thickness and isothermally at 70, 60 and 50 °C for both drying methods. Samples were blanched at 80 °C in hot water prior to drying. The color were measured using hunters method- a*, b*, L* and ΔE. The color were classified as under-dried, properly-dried, and over-dried and the machine learning algorithms used were namely Random Forest (RF), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) to trained on the dataset. The results indicated a superlative prediction accuracy of 100% by the RF model, followed by KNN with prediction accuracy of 98.3% and SVM with prediction accuracy of 95.7%. The research identified random forest model as the model with the best predictive capability among others, and therefore recommends its use in monitoring of the drying of potato slices given the conditions of this study.

Keywords: Random Forest; K-Nearest Neighbor; Support Vector Machine; Colour; Machine learning

1. Introduction

The increasing population of human beings and animals competing for scarce food supply for survival and production on daily basis as well as the increasing complexity of food production in recent times calls for advanced response to ensure quality control, safe food monitoring, and food production process enhancement. Several advancements are made nowadays in information processing which find application in numerous facets of human activities. Among the outstanding and versatile advancements is machine learning which involves vigorous efforts to teach modern information systems to learn hidden patterns that exist in the outcome of a particular domain of discourse through its generated data sets. Also, machine learning, in contrast to traditional problem-solving algorithms, does not program an extensive list of hard-coded instructions, but rather learns implicitly from data samples and generalizes to new cases based on their proximity to learned samples (instance-based learning) or trains a model with data to learn its parameters through optimization and makes predictions using new data in what is known as model-based learning (Deng et al., 2021). Among machine learning techniques and algorithms are decision trees, nearest neighbors, goal programming, inter-programming, genetic (or evolutionary) algorithms, linear probability model, neural networks, support vector machine, k-nearest neighbors, and random forest (Asongo et al., 2021).

Recent advancements in machine learning were explored, pointing out its applications in food quality, defect detection, visual inspection systems, ingredient optimization, nutritional assessment, food packaging, supply chain management in food industry, and industry 4.0 models (Liakos et al., 2025). The potentials of Artificial Intelligence and Machine Learning to revolutionize the food production processes had been explored in the area of defect detection, consistency

* Corresponding author: Mfrekemfon Godswill Akpan.

evaluation, intelligent food data analysis, real-time food monitoring, and pattern recognition (Chhetri, 2023). An extensive review of current trends in applications of machine learning models in agricultural product drying processes was presented (Fan et al., 2025). The review provided insight into the crucial stages in developing machine learning models consisting of data collection during drying, data pre-processing, model tuning, and model evaluation. The review also examined the principles and optimization strategies for supervised learning algorithms such as recurrent neural networks and long short-term memory networks.

2. Methodology

2.1. Sample Dataset Analysis

Fresh orange-fleshed sweet potatoes were purchased from Akpan-Andem market, a popular daily market in Uyo, Akwa Ibom state. They were selected based on uniformity in size, maturity, and absence of physical defects to minimize variability in experimental results. The samples were thoroughly washed with running water to remove surface dirt, peeled using a stainless-steel knife to prevent enzymatic browning, sliced into uniform thicknesses (typically 3mm) using a mechanical slicer to ensure consistency in drying behavior and then measured (thickness) using Vernier Calliper. The samples were blanched in hot (80 °C)water in a thermostatic bath for 4 mins to avoid enzymatic browning.

Materials used in the experiment included mechanical slicer (Royal Industry, 0–6 mm adjustable), stainless-steel knife for peeling, 30x20 Polypropylene trays, 5l airtight plastic containers

Hot-air dryer , vacuum dryer (Nabertherm VDL 115, Germany; temperature range 20–200 °C), digital balance (0.01g precision), Konica-Minolta CR-400 colorimeter, vernier calliper (0.01 mm), stopwatch, camera (16 MP), Microsoft Excel, SPSS v.27 and Matlab R2025b.

2.2. Drying Experiments

The drying experiments were carried out in both Vacuum Dryer and Hot air dryer. The dryers were allowed to run empty for 30 minutes to allow the thermal equilibrium before samples were set in them. After the samples were allowed in, the readings of colour changes and mass loss were recorded at 30 minutes interval till there were no significant mass loss. Drying was done isothermally at three different temperatures for both drying methods. The temperatures were 50, 60, and 70 °C. The colour were reported in hunters scale of a*, b* and L* where a* represents red to green chromaticity; b* represents blue to yellow chromaticity and L* represents lightness spanning from black to white. The total colour change was calculated using Equation 1. The moisture content was reported in wet basis.

$$\Delta E = \sqrt{(L_t^* - L_0^*)^2 + (a_t^* - a_0^*)^2 + (b_t^* - b_0^*)^2} \quad \text{Equation 1}$$

Where

ΔE is the total color change

L^*_t is the value of L at time t

L^*_o is the value of L initially/ penultimate time

a^*_t is the value of a* at time t

a^*_0 is the value of a* initially/ penultimate time

b^*_t is the value of b* at time t

b^*_0 is the value of b* initially/ penultimate time

2.3. Data Preprocessing

Before any machine learning activity can take place, the dataset must be rendered in a form suitable for the exercise in a process known as data preprocessing. In the data preprocessing phase, the following activities took place:

- S/N attribute was removed as it had no significant contribution to the training.

- The textual drying method attribute was converted to numerical form by representing Hot Air Drying (HAD) as 1 and Vacuum Drying (VD) as 2.
- The ^{o}C unit in the temperature attribute was removed from each of the record.
- Time values expressed in mixed minutes and hours unit was converted and expressed in minutes.
- The drying state was in textual form such as under-dried, properly-dried, and over-dried, which could pose problems to proper training. It was eventually coded as: under-dried = 1; properly-dried = 2; over-dried = 3.
- The drying state attribute was represented as a categorical data type.
- A duplicate Time attribute was removed.

Dataset analysis revealed that the initially generated data had a serious challenge of class imbalance as it was heavily biased in favour of the properly dried potato slices, therefore having a very high probability of generating obvious and misleading results by machine learning classifiers. A dataset is considered imbalanced if the existing classes are not approximately equally represented. Typically, machine learning algorithms performance is evaluated using overall predictive Accuracy, which is however not appropriate when the dataset is not balanced (Rachmatullah, 2022). In the given dataset, the "under-dried" and "over-dried" classes had very few data samples compared to the "properly-dried" class. To address this challenge, Synthetic Minority Oversampling Technique (SMOTE) presented in (Chawla et al., 2002) was used to generate synthetic minority samples by interpolating between existing ones, thereby filling the gaps and balancing the dataset without simply duplicating values. SMOTE is widely used in machine learning to improve classification accuracy on datasets that are imbalanced and has been successfully applied in agriculture, engineering, and biomedical studies to strengthen minority class detection.

The SMOTE algorithm shown as pseudocode is presented as follows (Chawla et al., 2002):

Algorithm *SMOTE*(*T*, *N*, *k*)

Input: Number of minority class samples *T*; Amount of SMOTE *N*%; Number of nearest

neighbors *k*

Output: (*N*/100) * *T* synthetic minority class samples

1 (*If N is less than 100%, randomize the minority class samples as only a randompercent of them will be SMOTEd*)

2 **if** *N* < 100

3. **then** Randomize the *T* minority class samples

4. *T* = (*N*/100) * *T*

5. *N* = 100

6. **endif**

7. *N* = (*int*)(*N*/100) (*The amount of SMOTE is assumed to be in integral multiples of100.*)

8. *k* = Number of nearest neighbors

9. *numattrs* = Number of attributes

10. *Sample*[ ][ ]: array for original minority class samples

11. *newindex*: keeps a count of number of synthetic samples generated, initialized to 0

12. *Synthetic*[ ][ ]: array for synthetic samples (*Compute* k *nearest neighbors for each minority class sample only.*)

13. **for** *i* ← 1 **to** T

14. Compute k nearest neighbors for i, and save the indices in the *nnarray*

15. Populate(*N*, *i*, *nnarray*)

16. **endfor**

Populate(N, i, *nnarray*) (*Function to generate the synthetic samples.*)

17. **while** $N \neq 0$

18. Choose a random number between 1 and *k*, call it *nn*. This step chooses one of the *k* nearest neighbors of *i*.

19. **for** attr ← 1 to *numattrs*

20. Compute: *dif* = *Sample*[*nnarray*[*nn*]][*attr*] − *Sample*[*i*][*attr*]

21. Compute: *gap* = random number between 0 and 1

22. *Synthetic*[*newindex*][*attr*] = *Sample*[*i*][*attr*] + *gap* * *dif*

23. **endfor**

24. *newindex*++

25. *N* = *N* − 1

26. **endwhile**

27. **return** (*End of Populate.*)

End of Pseudo-Code.

After applying SMOTE, each class had approximately equal representations, making the dataset more robust for ML training and reducing the risk of biased predictions.

Fifteen (15) attributes were used for the machine learning training as shown in Table 1.

Table 1 Attributes Description

Attribute	**Description**
Temperature	Drying temperature in ^{o}C
Weight_Init	Initial weight of sample in gram
Weight_Final	Final weight of sample in gram
L_before	Lightness of sample before drying
a_before	Red-green variation before drying
b_before	Blue-yellow variation before drying
L_after	Lightness of sample after drying
a_after	Red-green variation after drying
b_after	Blue-yellow variation after drying
DeltaE	Color difference
Time_min	Drying time in minutes

MC_percent	Moisture content
MR	Moisture ratio
Drying_Method	Method of drying: Hot Air Drying (HAD), Vacuum Drying (VD)
Drying_State	The state of drying: under-dried = 1, Properly-dried = 2, Over-dried = 3

A total of 576 data items with class distribution of 191 for class 1, 192 for class 2 and 192 for class 3 in the dataset were used for training. The dataset was randomized before being presented to the machine learning algorithm as shown in Table 2.

Table 2 Preprocessed Dataset

T	**W_i**	**W_f**	**L_i**	**a_i**	**b_i**	**L_f**	**a_f**	**b_f**	**ΔE**	**Time (min)**	**MC%**	**MR**	**DM**	**DS**
50	3.68	3.28	65.5	-3.72	11.05	52.05	-2.27	10.25	13.55	30	10.87	0.89	0	1
60	2.21	0.22	80.03	-2.95	25	50.55	-2.5	24.1	29.49	300	90.05	0.1	0	3
70	3.36	2.58	69.41	-0.25	15.36	43.34	-3.5	13.94	26.31	60	23.21	0.77	1	2
70	3.65	2.00	72.5	-4.82	26.99	43.16	-4.23	26.54	29.34	120	45.21	0.54	0	2
70	2.25	0.76	67.19	-0.75	14.88	49.47	-1.38	25.67	20.75	240	70.08	0.299	1	2
50	2.66	2.49	79.33	-0.238	17.13	32.78	-2.25	12.49	46.78	30	6.39	0.94	1	1
60	2.21	0.22	80.03	-2.95	25.00	50.55	-2.5	24.10	29.49	300	9.04	0.10	0	3
70	3.21	2.87	71.45	0.26	17.98	51.45	-2.87	15.01	20.46	30	10.59	0.89	1	1
60	5.06	4.71	69.49	-0.68	16.77	36.88	-1.96	12.03	32.98	30	6.92	0.93	1	1
50	4.11	3.92	81.18	-2.53	19.5	36.56	0.47	16.32	44.83	30	4.62	0.95	1	1
60	3.3	2.85	79.50	-6.11	15.12	30.00	-6.00	10.10	49.75	90	13.64	0.86	0	1
60	3.42	1.85	69.04	-0.15	15.43	49.71	-1.44	17.61	19.49	180	45.90	0.54	1	2
70	4.15	2.67	69.55	0.52	15.39	61.64	-1.2	18.11	8.39	90	35.66	0.64	1	2
50	4.44	4.02	65.56	-2.5	20.22	35.69	-6.66	15.66	30.50	120	9.45	0.91	0	1
60	5.06	4.71	69.49	-0.68	16.77	36.88	-1.96	12.03	32.98	30	6.92	0.93	1	1
60	2.21	0.22	80.03	-2.95	25.00	50.55	-2.50	24.10	29.50	300	90.05	0.10	0	3
50	2.17	1.27	84.24	-1.14	18.61	62.29	-0.07	19.71	22.00	300	41.47	0.59	1	2
60	2.21	0.22	80.03	-2.95	25.00	50.55	-2.50	24.10	29.50	300	90.04	0.10	0	3
70	2.8	2.02	70.02	-4.50	25.20	49.26	-7.57	22.59	21.19	60	27.85	0.72	0	2
60	3.52	3.11	65.75	-1.15	14.31	52.92	-1.44	15.48	12.88	60	11.64	0.89	1	1
60	2.21	0.22	80.03	-2.95	25.00	50.55	-2.50	24.1	29.50	300	90.05	0.10	0	3
60	2.21	0.22	80.03	-2.95	25.00	50.55	-2.50	24.1	29.50	300	90.05	0.10	0	3
60	4.46	3.24	66.86	-0.46	18.01	3705	-2.27	13.56	30.20	120	27.35	0.73	1	2
70	3.84	1.32	70.27	-5.33	15.17	48.70	-6.25	14.99	21.59	300	65.63	0.32	0	2

Drying state attribute constitutes the response variable (or target/output variable) all other attributes represent the input variables. 80% (460 samples) of the dataset was used for training while 20% (115 samples) was set aside for

testing. The drying state of the potato slices consists of three classes to be predicted by the machine learner classifier namely, class 1 (under-dried), class 2 (properly-dried), and class 3 (over-dried).

Three machine learning algorithms, namely Random Forest (RF), K-Nearest Neighbor (KNN), and Support Vector Machine (SVM) were trained on the dataset. The Random Forest (RF) classifier is a bagged ensemble learning method that constructs multiple decision trees during training and selects the mode of their predictions as the computed output for classification activities. It helps in effectively reducing overfitting, which is a common challenge with single decision trees.

3. Results and discussion

After the RF model training, the testing result showed an overall prediction accuracy of 100.0%, indicating a perfect performance of the bagged ensemble in classifying the drying state of the sweet potato slices shown by the diagonal values in the confusion matrix of Figure 1.

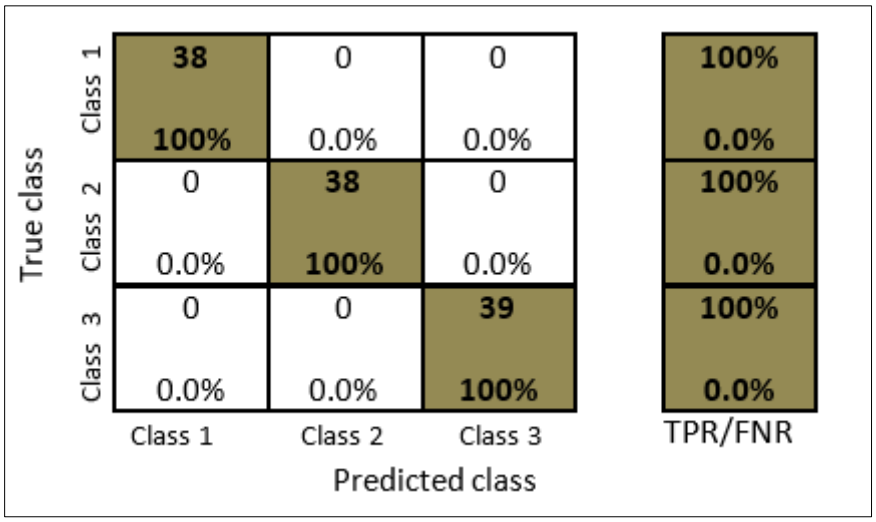

Figure 1 Random Forest Confusion Matrix

Results of RF modeling shows a perfect prediction where all the classes 1, 2 and 3 test samples were correctly predicted by the model. This indicates that all the 38 samples of class 1 (under-dried samples), 38 samples of class 2 (properly-dried samples) and 39 samples of class 3 (over-dried samples) were correctly predicted.

Two other powerful machine learning models were trained to provide a basis for comparison of models performance on the food sample. They are Support Vector Machine (SVM) and K-Nearest Neighbor (KNN).

Modeling with SVM shown in Figure 2 indicates an overall prediction accuracy of 95.7%.

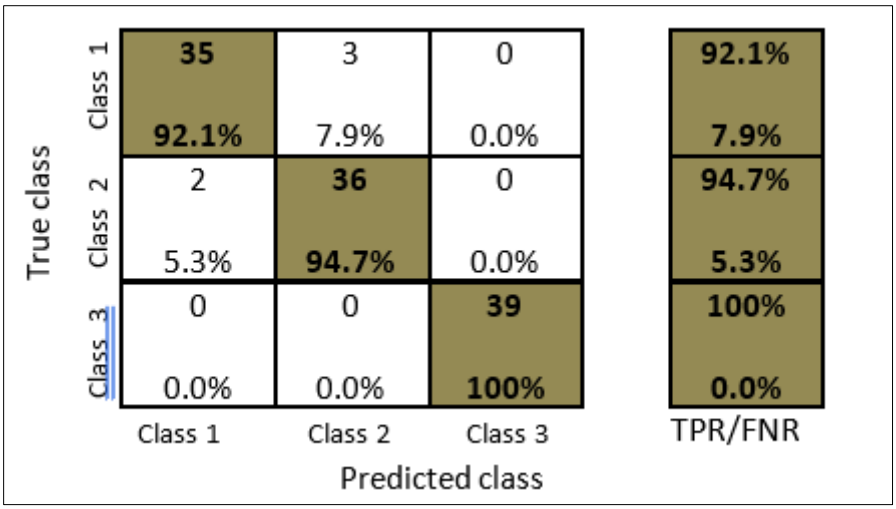

Figure 2 SVM Confusion Matrix

Here, only the over-dried sweet potato slice samples (class 3) with 39 samples were correctly (100%) predicted by the SVM model. The model mis-predicted 3 samples of class 1 (under-dried samples) out of 38 samples to be properly-dried. Moreover, 2 samples of class 2 (properly-dried samples) out of 38 samples were mis-predicted as under-dried samples.

K-Nearest Neighbor classifier modeling generated a model that has an overall prediction accuracy of 98.3% shown by the confusion matrix of Figure 3.

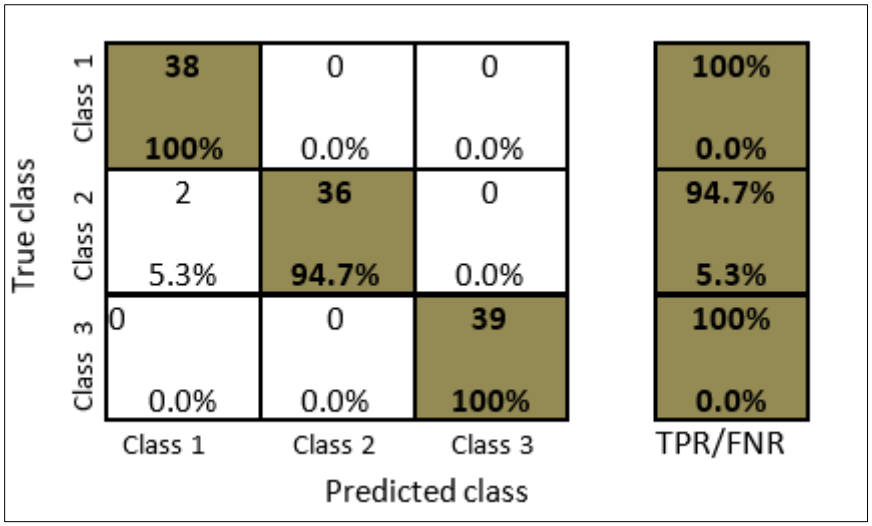

Figure 3 KNN Confusion Matrix

All the class 1 (under-dried) samples and all the class 3 (over-dried) samples were correctly predicted by the KNN model. The model however wrongly predicted 2 samples of the properly-dried sweet potato slices as being under-dried. The prediction performances of the three models are presented as shown in Figure 4.

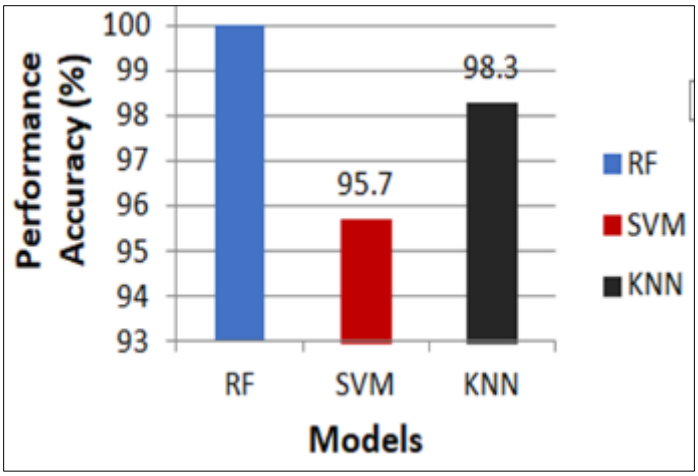

Figure 4 Model Performance Accuracy

The visualization of the performances of each model clearly indicates the RF model as the best performing model.

4. Conclusion

Competition for scarce food supply and the increasing complexity of food production processes calls for advanced responses to ensure quality control, safe food monitoring, and enhancement of food production. One of the current techniques of predicting the drying states of food samples is the application of machine learning algorithms to solve problems in the agricultural sector. Experiment was conducted on the drying processes of sweet potato slices for storage purpose among other purposes. The drying states of the samples are typically determined through manual inspection of the color variations. The drying states are under-dried, properly-dried, and over-dried. Three machine learning models namely Random Forest (RF), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN) were trained to carry out prediction of the drying states of the samples. The results indicated a superlative prediction accuracy of 100% by the RF model, followed by KNN with prediction accuracy of 98.3% and SVM with prediction accuracy of 95.7%. The research identified random forest model as the model with the best predictive capability among others, and therefore recommends its use in prediction of the drying states of sweet potato slices.

Compliance with ethical standards

Acknowledgement

This study was funded by TETFund Centre of Excellence in Computational in Computational Intelligence Research, University of Uyo, Uyo Nigeria.

Disclosure of conflict of interest

The authors do not have any conflict of interest to disclose.

Statement of informed consent.

All the authors have consented to publish this study at the International Journal of Science and Research Archive

References

[1] Asongo, A. I., Barma, M., & Muazu, H. G. (2021). Machine Learning Techniques, methods and Algorithms: Conceptual and Practical Insights. *International Journal of Engineering Research and Applications*, *11*(8), 55–64. https://doi.org/10.9790/9622-1108025564

[2] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

[3] Chhetri, K. B. (2023). Applications of Artificial Intelligence and Machine Learning in Food Quality Control and Safety Assessment. *Food Engineering Reviews*, *16*(1), 1–12. https://doi.org/10.1007/s12393-023-09363-1

[4] Deng, X., Cao, S., & Horn, A. L. (2021). Emerging Applications of Machine Learning in Food Safety. *Annual Review of Food Science and Technology*, *12*, 518–538. https://doi.org/doi.org/10.1146/annurev-food-071720-%2520024112

[5] Fan, L., Pei, Y., Zhang, L., Kong, J., & Xu, W. (2025). Applications of Machine Learning Models in Agricultural Product Drying: A Comprehensive Review of Advances, Challenges, and Prospects. *Food and Bioprocess Technology*, *18*, 10047–10085.

[6] Liakos, K. G., Athanasiadis, V., Bozinou, E., & Lalas, S. I. (2025). Machine Learning for Quality Control in the Food Industry: A Review. *Foods*, *14*(3424), 1–35. https://doi.org//doi.org/10.3390/foods14193424

[7] Rachmatullah, M. I. C. (2022). The Application of Repeated SMOTE for Multi Class Classification on Imbalanced Data. *Matrik: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer*, *22*(1), 13–24. https://doi.org/10.30812/matrik.v22i1.1803