Systematic review of artificial intelligence with near-infrared in blueberries

Liset Cayhualla Amaro, Sebastian Rau Reyes, María Acuña Meléndez, Christian Ovalle
Departamento de Ingeniería, Facultad de Ingeniería, Universidad Tecnológica del Perú, Lima, Perú

Article Info

Article history:

Received Jan 9, 2024
Revised Mar 3, 2024
Accepted Mar 21, 2024

Keywords:

Artificial intelligence
Blueberries
Chemometry
Fruit
Machine learning
Nir specters
Nondestructive evaluation

ABSTRACT

The fruit quality has a direct impact on how the fruit looks and how tasty the fruit is. The correct use of tools to determine fruit quality is essential to offer the best product for the final consumer. This study has used the preferred reporting items for systematic reviews and meta-analyses (PRISMA) methodology. The study objective was elaborate a systematic literature review (SLR) about research of the application of techniques based on artificial intelligence to analyze indicators obtained by near infrared spectroscopy (NIRS) and chemometrics to determine the quality of fruits, including blueberries. The most frequently addressed indicator is the soluble solids concentration (SSC) which was used in several studies with techniques such as support vector machines (SVM) and convolutional neural networks (CNN). According to the results obtained, it is possible to use these techniques to predict blueberry quality indicators. There was an acceptable performance and high accuracy of these models. However, future research could cover other techniques and help to provide better quality control of products in food industries.

Corresponding Author:

Christian Ovalle
Departamento de Ingeniería, Facultad de Ingeniería, Universidad Tecnológica del Perú
Lima, Perú
Email: dovalle@utp.edu.pe

1. INTRODUCTION

Blueberries are food with high nutritional value, and they have benefits related to health to impact on quality of people´s lives. Blueberries also are tasty for most people. They have cultivation qualities that are useful for rapidly growing and that's why their popularity has been spreading around the world [1]. As blueberries became popular, people also demanded an acceptable quality for it. Blueberry production is also increasing year after year, so the standard for picking is the fruit in a specified maturity time [2].

Blueberries' ripeness affects how they taste and how long they can be stored without getting rot. Some quality testing combines ripening parameters and other characteristics based on pre-harvest factors. In a common process, the blueberry quality test can evaluate some key characteristics like soluble solids concentration (SSC) to determine a correct harvest maturity time, however this is done destructively by discarding several blueberry samples [3]. Sugar is the main component of fruit quality and flavor. It also affects the aromatic compounds and color pigments of blueberries [4].

This study investigated how near infrared spectroscopy (NIRS) with other technologies like machine learning and techniques like chemometrics could be used to improve quality evaluation of blueberry fruit in a national context. NIRS has been mostly intended for use in the agri-food industry and it is employed at different stages of a grower's supply chain [5]. NIRS has been used as a rapid, non-destructive screening tool to determine certain fruit attributes for postharvest quality evaluation [6].

In combination with chemometrics techniques, it has shown great potential for the determination of physical and chemical parameters such as total soluble solids (TSS) and total titratable acidity (TA). This technique allows for determining physicochemical properties related to various biochemical changes such as fruit ripening. Furthermore, large amounts of data may be available that require multivariate statistical approaches to identify the most relevant information from large data sets and create prediction and classification models with practical applications [4].

In some research, we found uses of NIRS in combination with chemometrics for the evaluation of SSC in different fruits to optimize the time of quality evaluation and increase its efficiency. According to Cerezo *et al.* [7], NIRS and principal component analysis (PCA) are used to determine the quality of olive oil based on the SSC, which determines a higher value for the prediction and classification of maturity. According to Santos *et al.* [8], it explored the use of NIRS and partial least squares regression (PLSR) to evaluate the maturity of citrus fruits. It also used preprocessing techniques such as multiplicative scatter correction (MSC) and standard normal variate (SNV) that achieved better precision of the SSCs of the fruits. Montecchiarini *et al.* [9] proposed machine learning models such as convolutional neural network (CNN) and feedforward neural network (FNN) in combination with NIRS and PCA for the evaluation of SSC in O'Neal and Emerald variety blueberries with an accuracy of 98%. According to Rungpichayapichet *et al.* [10], it used partial least squares (PLS) regression as a prediction model for mango SSC with a precision of 87%, the result was very low due to the time in which the harvests were carried out, which determined a relationship that was not so appropriate. For predictions, Ferrara *et al.* [11] developed predictive models for the evaluation of the quality of the grapes located in the vineyards, which determines the exact time that should be harvested, however, some limitations were found for collecting data directly in the vineyards, such as the shade that disperse the data. Ditcharoen *et al.* [12] proposes models such as support vector machine (SVM) and k-nearest neighbors (KNN) for the prediction of SSCs of durian fruit, different techniques such as MSC and SNV were used to calibrate the spectral data, which helps to improve model training which had an approximate accuracy of 89%. However, it is recommended to have more test data to improve the performance of the models.

In summary, our investigation suggests that a well-performing blueberry SSC detection model can help farmers and laboratories to quickly detect the ripeness of blueberries more accurately. This allows them to identify the optimal blueberry harvest time and reduce the current quality evaluation that requires a lot of time and manpower. That is why in the present review, we explored different artificial intelligence models in combination with NIRS and chemometric techniques for SSC detection. Likewise, various research will help to deepen the relationship between blueberry sugar content and its ripeness to facilitate this detection.

2. METHOD

2.1. Systematic search strategy

In this research the PICO method was used. Its name is the initials of the words P (population), I (intervention), C (comparison), and O (outcome). This method allows the formulation of strategic questions to carry out more precise and effective searches for articles [13]. In addition, the formulation of the PICO question and its components that helped to formulate the research question was carried out and can be seen in Table 1. Additionally, a disaggregated PICO table was made where the keywords in each component were identified and thus generate the search equation that will help to obtain more accurate results; this can be seen in Table 2.

2.2. Search equation

Next, searches were performed in two databases focused on research, which are Scopus and PubMed, where the equation was formulated based on the keywords defined in Table 2. Additionally, other filters were added, such as that the publications should be between 2019 and 2023 and that the documents should be systematic reviews and articles. As a result, a total of 265 investigations were obtained for Scopus and a total of 8 investigations were obtained from PubMed. The following is described in Table 3.

Table 1. PICO question and its components

Question PICO	Components
– RQ: How would the use of a machine learning model based on NIRS and chemometrics improve the evaluation of the internal quality of blueberries compared to traditional destructive evaluation?	– RQ1: What kind of blueberry quality indicators have been predicted using machine learning methods? – RQ2: What are the different machine learning techniques that have been applied to predict blueberry quality? – RQ3: What are the chemometric techniques used in blueberry quality prediction? – RQ4: How have you carried out the performance metrics of machine learning models to predict the internal quality of blueberries?

Table 2. Disaggregated PICO table

		Keyword	Equation syntax
P: Problem/ Population	Blueberries for quality evaluation	Postharvest, fruits, berries	Postharvest OR fruits OR berries
I: Intervention	Machine learning model based on NIRS and chemometrics	Machine learning, NIR, near infrared spectroscopy, chemometry, artificial intelligence	"machine learning" OR "nir" OR "near infrared spectroscopy" OR "chemometry" OR "artificial intelligence".
C: Comparison	Traditional destructive evaluation	Maturity, soluble solids, titratable acidity, destructive method, traditional".	maturity OR "soluble solids" OR "titratable acidity" OR "destructive method" OR "traditional".
O: Results	Improved internal quality assessment	Quality evaluation, results, non-destructive, evaluation, evaluation	"quality evaluation" OR "results" OR "non-destructive" OR "evaluation".

Table 3. Database search

Database	Equation	Results obtained
SCOPUS	TITLE-ABS-KEY (postharvest OR fruits OR berries AND "machine learning" OR "nir" OR "near infrared spectroscopy" OR "chemometry" OR "artificial intelligence" AND maturity OR "soluble solids" OR "titratable acidity" OR " destructive method" OR " traditional" AND "quality evaluation" OR "results" OR "non-destructive" OR "evaluation") AND PUBYEAR > 2018 AND PUBYEAR < 2025 AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "re")) AND (LIMIT-TO (OA, "all"))	265
PUBMED	(postharvest OR fruits OR berries) AND ("machine learning" OR "nir" OR "near infrared spectroscopy" OR "chemometry" OR "artificial intelligence") AND (maturity OR "soluble solids" OR "titratable acidity" OR " destructive method" OR " traditional") AND ("quality evaluation" OR "results" OR "non-destructive" OR "evaluation")	8

2.3. Inclusion and exclusion criteria

Inclusion and exclusion criteria were then developed based on the articles found in the databases. Important topics that should be included in the papers were identified to help make a better selection for the research. The criteria are defined in Table 4.

Table 4. Inclusion and exclusion criteria

Inclusion criteria	Exclusion criteria
CI1: Studies should include predictive models of quality parameters. CI2: Studies should include the implementation of chemometrics and machine learning. CI3: Studies should include the use of Nir spectra in fruits. CI4: Studies should include machine learning or deep learning approaches to classify a fruit. CI5: Studies should include methods for information extraction to train machine learning models. CI6: Studies should include chemometric techniques.	CE1: Documents that do not use NIRS for information capture. CE2: Documents that do not implement chemometrics. CE3: Papers that do not mention machine learning techniques. SC4: The documents that do not speak of the quality of the fruits. CE5: Studies that do not deal with machine learning or artificial intelligence or deep learning. CE6: Documents that are prior to 2019. CE7: Publications that are different from English or Spanish.

2.4. Selection process

For the selection of articles, the PRISMA flowchart was used to help identify research articles that best fit the topic of systematic reviews by filtering by exclusion and inclusion criteria [14]. The first step was eliminating duplicate articles, then validating compliance with the inclusion and exclusion criteria, and finally obtaining the articles relevant to the research. Next in Figure 1, a review of 273 documents selected from the search in the databases was carried out, where 265 documents were selected from Scopus and 8 documents from PubMed. Of the total number of articles, the review of the abstract as well as the title of each article was carried out and a total of 130 were eliminated for not including the topics of NIRS and chemometrics, in addition to not including mention of artificial intelligence topics such as predictive models, machine learning or deep learning. 143 documents remained with the defined topics. Then 8 articles have been eliminated because they are not fully available for a complete review. After reviewing the remaining 135 articles, 90 articles have been excluded for focusing only on spectral imaging and portable data acquisition devices; 1 article for addressing the use of artificial intelligence as the main topic; 3 articles for focusing on FT-NIR and 1 article for addressing the use of nitrogen isotope radiation. 40 articles met the criteria for inclusion. The PRISMA flow chart prepared for the research is shown in Figure 1.

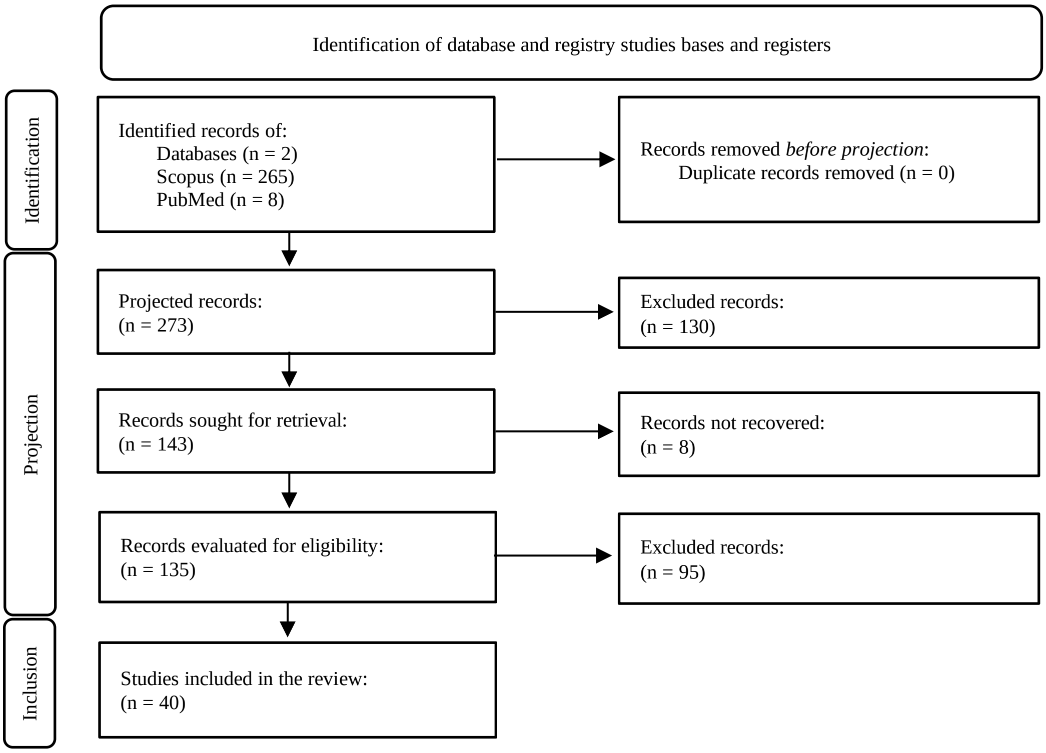


Figure 1. PRISMA flow diagram

3. RESULTS AND DISCUSSION

3.1. Bibliometric analysis

Bibliometric analysis is very valuable since it facilitates the interpretation of large amounts of data and provides a clearer insight into novel ideas for research [15]. After the application of the document selection process, 40 studies were obtained for the systematic review. For the elaboration of Figure 2, the systematic review main terms were "near infrared spectroscopy", "analysis", "fruits", and "machine learning" included in the titles, keywords or abstracts of these papers. Specialized data analysis software known as VOSviewer was used and according to Rodriguez *et al.* [16] it serves to build bibliometric networks based on data downloaded from bibliographic databases and thus determine the relevance and relationship between the key components of the review.

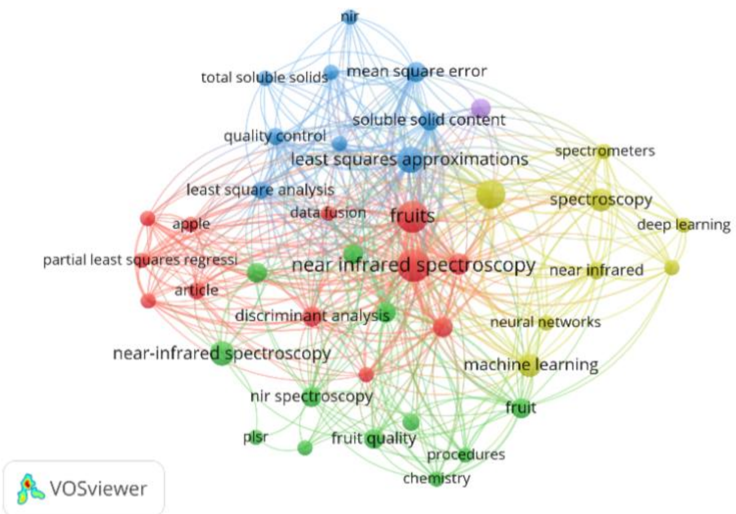


Figure 2. Analysis of connections between research keywords

Figure 3 shows the representation of the proportion of the main terms that indicate the frequency of publication by year of the articles. In addition, the relationship by color can be seen, with "deep learning" and "chemistry" as recent terms in research published in the last half of 2022. The density of the key elements can be appreciated where two of the most predominant terms within the research topic can be visualized which are "near infrared spectroscopy" and "fruits" referring to the articles that include these topics published in the last years.

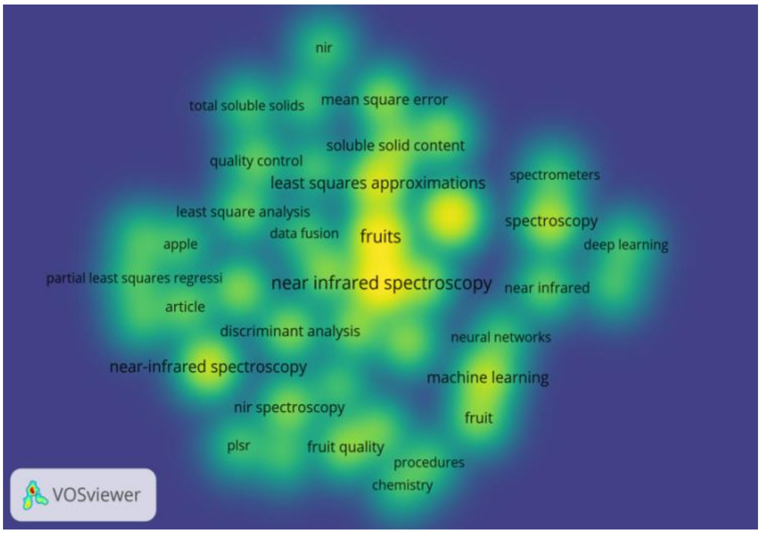


Figure 3. Keyword density analysis of research keywords

3.2. Year of publication

The selected studies were published between 2019 and 2023. Figure 4 shows the number of studies published within the selected timeline. Overall, the oldest study reviewed was published in 2019, while in 2022 most of the articles were published. Of those reviewed, four articles were published in 2019, five articles in 2020, three articles in 2021, fifteen articles in 2022, and thirteen articles in 2023. Based on the results, it is observed that the latest studies touch on topics related to applications of artificial intelligence as one of the approaches to improve the prediction of results based on NIRS.

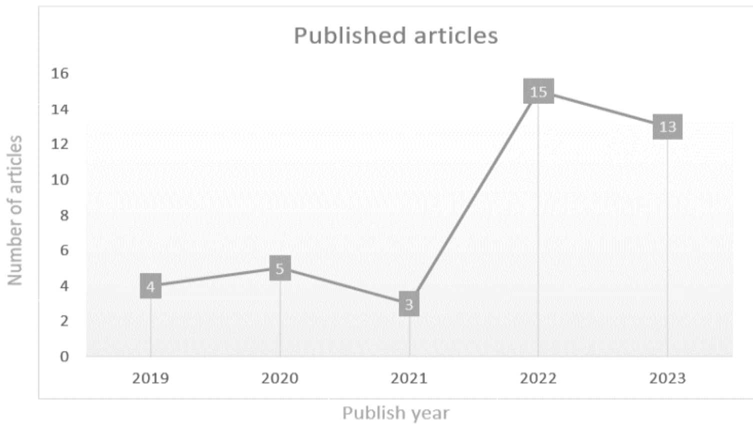


Figure 4. Number of documents by year of publication

3.3. Blueberry quality indicators

The research questions are discussed in this section. The first research question addressed is: (RQ1) What type of blueberry quality indicators have been predicted using machine learning methods? Based on the articles reviewed, various internal quality indicators of fruits have been found to be important in determining ripeness. These can be distinct compounds present inside the fruit that can be detected by spectroscopic means or by destruction of the fruit [17]–[21]. Some internal indicators such as sugar content, acid-sugar ratio, firmness, texture, and vitamin C are obtained by traditional destructive and non-destructive methods to be used to determine the quality of the fruit [22]–[25]. Likewise, according to Pornchaloempong *et al.* [26], NIRS was used to predict the SSC of mangoes, which indicates the sweetness of the product and determines whether it meets marketing standards. Furthermore, according to Zhang *et al.* [23], soluble solids, firmness, and acidity are established as important internal quality characteristics. Also, SSC is an indicator that is more linked to consumer perception of fruit quality [27], [28]. Table 5 shows that the SSC indicator is the most used and evaluated in research because it is a value that helps to determine the content of internal sugars, as well as other soluble compounds such as certain minerals, which are key to determine the quality and ripeness of fruits [25], [27], [29]–[33]. Also, according to Cai *et al.* [34], the prediction of the moisture indicator helps to know the level of dehydration present in the fruits and this in turn was useful to determine how much the fruit was affected in its texture and flavor characteristics which are of fundamental importance for the final consumer. Similarly, the TA indicator is useful to determine the amount of citric acid contained in the fruit, which is an important factor that maintains the balance between sweetness and acidity [19], [20], [26], [30], [35], [36]. In addition, other articles talk about how detect the indicator phenolic acids to determine if the fruit is susceptible to spoilage [17], [24]. This helps to protect fruits because the higher the phenolic acid indicator, the longer the preservation and the longer the shelf life of the fruit. Vitamin C is also detected as an indicator that, Ye *et al.* [18] determines whether the fruit can maintain adequate freshness for a longer time depending on its quantity and helps to preserve it. Finally, the dry matter (DM) amount present in the fruit is also taken as an indicator [37]. This indicator determines whether the fruit can maintain an adequate freshness for a longer time. A range between 10% to 30% of the recommended presence in the fruit is also useful to determine its maturity [38], [39].

Table 5. Indicators for quality assessment

#	Quality indicator	Unit of measurement	Quantity	Articles
1	Soluble solids concentration	Brix	33	[11], [18]–[27], [28]–[31], [33], [35], [39]–[46], [47]–[52]
2	Amount of humidity (CH)	%	2	[34], [48]
3	Titratable acidity	%	10	[11], [19], [20], [24], [26], [29], [35], [36], [47]
4	Phenolic acids (FT)	%	2	[18], [20]
5	Dry matter	%	6	[23], [38], [48], [52], [53]
6	Vitamin C (VC)	%	2	[18], [19]

3.4. Most used machine learning techniques

The second research question addressed was the following: (RQ2) What are the different machine learning techniques that have been applied to predict blueberry quality? Based on the analysis of the articles, some machine learning techniques were identified from those that have been associated with 3 of the internal quality indicators that are the most important for blueberry quality. Table 6 shows first the SSC prediction indicator, they implemented a SVM algorithm to predict the Brix of fruits such as pears, oranges, oranges, and pears [24], [29], [40], [43], [44]. Then were 4 articles that implement the CNN for SSC prediction [19], [28], [30], [39]. Least-squares support vector machine (LS-SVM) based models were also used for fruit maturity classification [27], [31], [33], [38]. Also detailed are models such as graph convolutional network (GCN), KNN, and support vector machines regression (SVMR) which are models that help to improve regression and classification problems by handling nonlinear inputs as linear outputs [19], [25], [29], [35], [42]. Then there is the TA concentration prediction indicator, which in the articles was identified as using the ANN algorithm, which is a neural network composed of input layers where the spectra data are entered. There is the hidden layer where the calculation and processing are performed and the output layer that shows the % of TA [26], [37]. In addition, Basile *et al.* [54] indicates that increasing the hidden layers will result in an erroneous prediction, also the training should be divided into data to train the model at 80% and data to test the model at 20%. Also, SVM was used for the classification of sweetness of oranges, which were classified into three classes: sweet, mixed (sweet and sour), and sour [35]. Additionally, models such as GCN and CNN are also detailed in [19] the combination of these models shows an accurate performance in predicting the internal quality parameters of mandarins since a rapid and non-destructive evaluation is performed to obtain the data.

Finally, we have the DM prediction indicator, which was identified in the articles as using a model based on artificial neural networks (ANN), which is a subset of machine learning that helps to process data and solve problems [54]. Moreover, according to Puttipipatkajorn *et al.* [37] for ML algorithms to have better results, spectral pretreatment with calibration models such as SNV, modified sine cosine algorithm (MSCA) should be performed. Also, Subedi and Walsh [52] mentioned that the ANN model has used a single hidden layer architecture, which helped the prediction of attributes such as MS and obtained results like those achieved by models such as PLSR.

Table 6. Most used machine learning techniques by quality indicator

#	Quality indicator	ML techniques	Quantity	Articles
1	SSC prediction	SVM	5	[29], [35], [43], [44]
		CNN	4	[19], [28], [39]
		LS-SVM	4	[19], [27], [31]
		GCN	1	[33]
		KNN	2	[29], [35]
		SVMR	2	[25], [42]
		Linear regression model (LRM)	1	[40]
		Multilayer perceptron (MLP)	1	[40]
		Linear discriminant analysis (LDA)	2	[29], [35]
2	Prediction of TA concentration	SVM	2	[29], [35]
		GCN	1	[19]
		CNN	1	[19]
		KNN	2	[29], [35]
		LDA	2	[29], [35]
		ANN	3	[37], [54]
3	Dry matter prediction	ANN	4	[37], [53], [54]

3.5. Chemometric techniques for predicting blueberry quality

The third research question that was addressed is the following: (RQ3) What are the chemometrics techniques used in the prediction of blueberry quality? Based on the analysis of the articles, some chemometrics based techniques have been identified to predict key indicators for determining fruit quality through analysis. As detailed in Table 7, the techniques with the highest mention in the research are shown grouped by the indicator to be predicted. Starting with the prediction of SSC, which is obtained by employing PLSR among other techniques such as the application of PCA to search for differences in spectral data and develop calibration models to be used with data obtained by NIRS [17], [22], [35], [41]. Guo *et al.* [50] mentioned that it is possible to apply regression with multiple latent variables (PLS2R) to work the NIRS data in a multilinear regression. In addition, it is denoted that the presence of environmental noise or external factors can generate nonlinear spectra that could alter the expected results so that these anomalies should be treated with classification techniques such as SVM [11], [47], [49]. In addition, according to Ye *et al.* [18] details that partial least squares discriminant analysis (PLS-DA) serves for qualitative analysis that can be used to elaborate the classification of fruit maturity characteristics and attributes. Furthermore, PLS is a multivariate method that can also be used for spectral data processing and can also analyze redundant variables to obtain more detailed information [21], [33], [40], [55]. Finally, for the prediction of TA and DM concentration, PLSR and PLS methods were used to model the spectral data obtained by NIRS to identify the ripening stages of the fruits and establish the relationship between the predictor variables [26], [38]. Also, according to Zeb *et al.* [35], PCA is used as a method to reduce the set of data that can be displayed by the spectra and reduce them to fewer variables to obtain a better analysis of them.

Table 7. Most used chemometric techniques by quality indicator

#	Quality indicator	Chemometric techniques	Quantity	Articles
1	Soluble solids concentration prediction	Partial least squares regression	8	[11], [17], [22], [28], [35], [41], [47], [48]
		Partial least squares discriminant analysis	1	[18]
		Principal component analysis	6	[11], [18], [27], [35], [49]
		Partial least squares	7	[29], [30], [33], [40], [46], [55]
		Sparse partial least squares regression (SPLSR) and sparse partial least squares multinomial regression (SPRMR)	1	[41]
		Interval partial least squares (iPLS)	1	[44]
		Successive projections algorithm (SPA)	1	[50]
2	Prediction of TA concentration	Partial least squares regression	1	[26]
		Partial least squares	1	[29]
3	Dry matter concentration prediction	Partial least squares	1	[38]
		Partial least squares regression	3	[48], [52], [53]

3.6. Performance of machine learning methods

The fourth research question addressed: (RQ4) How have the performance metrics of machine learning models been conducted to predict the internal quality of blueberries? Based on the analysis of the articles, performance metrics that assess the accuracy or predictive ability of a machine learning model have been identified. As detailed in Table 8, the different metrics adopted are shown. According to Zeb *et al.* [35] the accuracy metric was used to evaluate the accuracy of the joint SVM+KNN model which was used to predict the percentage of SSC in oranges, obtaining an accuracy of 81.03% for the classification. Likewise, according to Pourdarbani *et al.* [36], the accuracy metric was used to evaluate the CNN model, which obtained an accuracy of 100% for the detection of pH of apples and the coefficient of determination was 0.86 which determined the variability of the data for predicting pH. Furthermore, according to Wu *et al.* [19] a GCN-LSTM-AT model was used to predict SSCs in mandarins, which achieved the best performance compared to other models.

Table 8. Model performance evaluation metrics by quality indicator

#	Quality indicator	Selected model	Performance metrics	Result	Article
1	SSC	SVM and KNN	Accuracy (%)	81.03	[35]
2	SSC	GCN-LSTM-AT	RMSECV	0.1430	[19]
			R^2	0.988	
			MAE	0.119	
3	pH	ANN	Accuracy (%)	100	[36]
			R^2	0.86	
4	TA	ANN	Accuracy (%)	99.2	[36]
			R^2	0.86	
5	SSC	CNN	Accuracy (%)	97.1	[39]
6	SSC	LS-SVM	Root mean square error of prediction (RMSEP)	0.32	[33]
7	SSC	SVMR	R^2	0.95	[42]
			RMSE	1.83	
8	SSC	ANN	RMSE	0.52	[54]
			R^2	0.82	
9	SSC	CNN	Accuracy (%)	98.9	[39]
			R^2	0.71	
10	SSC	CNN	R^2	0.8580	[28]
			RMSE	0.4276	
11	SSC	SVM	Accuracy (%)	85	[43]
12	SSC	LDA	Accuracy (%)	91.43	[29]
13	SSC	CNN	R^2	0.87	[30]
			RMSE	1.76	
14	SSC	LDA+SVM	Accuracy (%)	97.44	[44]
15	SSC	LS-SVM	R^2	0.973	[31]
16	SSC	SVM-R	RMSEP	1.6867	[25]
17	MS	ANN	RMSEP	0.89	[53]
18	SSC	BPNN	R^2	0.8872	[32]
			RMSE	0.4709	
19	Degree of bruising	LS-SVM+SPA	Accuracy (%)	97.3	[27]

The metrics considered to evaluate the model were root mean square error of cross-validation (RMSECV) which obtained (0.1430∘Brix), R^2 (0.988), and mean absolute error (MAE) (0.119∘Brix). Likewise, according to Galal *et al.* [42], the metrics coefficient of determination (R^2) and root mean squared error (RMSE) were used to evaluate the effectiveness of the SVMR models for the evaluation of bananas and ANN for the evaluation of grapes, which obtained a better performance for the prediction of soluble solids, for the SVMR model a V (0.95), RMSE (1.83) was obtained and the ANN model obtained a R^2 (0.95), RMSE (1.83). In addition, Escárate *et al.* [39] the estimation of soluble solids in stone fruits is performed, for this purpose a model based on CNN was performed, which obtained a performance of accuracy (98.9%) and R^2 (0.71) compared to the other models. Likewise, according to Xu *et al.* [28] performance of the CNN-based model for the prediction of SSC in oranges was evaluated and the results were R^2 (0.8580) and RMSE (0.4276). In addition, according to Lazim *et al.* [43] the SVM-based model was evaluated for the prediction of SSC and classification of watermelon maturity level, obtaining a result of 85% accuracy. The LDA model was also evaluated for mango maturity level prediction, where the Accuracy performance metric was used and obtained a result of 91.43% [29]. Likewise, according to Kalopesa *et al.* [30] the CNN model is used for the estimation of sugar content in grapes, having an evaluation of the model an R^2 (0.87), RMSE (1.76). Research by Lamptey *et al.* [44] the LDA-SVM model was evaluated for the classification of mangoes where an accuracy of 97.44% was obtained, however, if identification techniques such as mean centering (MC), SNV, first derivative (FD), and second derivative (SD) are used, a higher percentage of accuracy could be achieved. Finally, Shao *et al.* [27] the combination of LS-SVM models was used for the classification of cherries based on the degree of bruising, which resulted in an accuracy of 97%. The following reviews have shown that several types of AI techniques have been applied for prediction and different metrics for evaluating the performance of the models, which demonstrated a correct result in the precision of the different evaluated chemical aspects of a fruit.

4. CONCLUSION

This study presented a systematic review of the literature about how the application of NIRS with artificial intelligence in conjunction with chemometrics allows the analysis and evaluation of the quality of blueberries and other fruits. The quality of these fruits can vary due to various human and environmental factors from the production stage to the distribution and consumption stage. The importance of understanding quality will improve consumer satisfaction and allow fruits to be preserved in their optimal state of consumption. Recent observations suggest that the way to analysis of the quality of the fruit is done through techniques such as NIRS, chemometrics, and artificial intelligence models that seek to improve the traditional ways of internal evaluation of the quality of blueberries and find relationships between the analyzed spectra and the ° Brix of blueberries. Likewise, our research provides conclusive evidence that the implementation of predictive models for the prediction of SSC in blueberries and other types of fruits achieved a good result in quality evaluation. This study is expected to contribute knowledge for future research work on how the application of artificial intelligence, NIRS, and chemometrics can provide better ways to determine the quality of fruits, including blueberries.

REFERENCES

[1] Y. Liu *et al.*, "'Is this blueberry ripe?': a blueberry ripeness detection algorithm for use on picking robots," *Frontiers in Plant Science*, vol. 14, 2023, doi: 10.3389/fpls.2023.1198650.

[2] W. Yang, X. Ma, and H. An, "Blueberry ripeness detection model based on enhanced detail feature and content-aware reassembly," *Agronomy*, vol. 13, no. 6, 2023, doi: 10.3390/agronomy13061613.

[3] Y. Bai, Y. Fang, B. Zhang, and S. Fan, "Model robustness in estimation of blueberry SSC using NIRS," *Comput Electron Agric*, vol. 198, 2022, doi: 10.1016/j.compag.2022.107073.

[4] H. Yang *et al.*, "Transcriptomic and metabolomic profiling reveals the variations in carbohydrate metabolism between two blueberry cultivars," *International Journal of Molecular Sciences*, vol. 25, no. 1, 2024, doi: 10.3390/ijms25010293.

[5] M. V. -Castellote, D. P. -Marín, I. T. -Rodríguez, J. M. M. -Rojas, J. L. O. -Díaz, and M. T. Sánchez, "Green, multivariate approach for obtaining a fingerprint of quality of watermelons at supermarket level using near infrared spectroscopy," *LWT*, vol. 182, 2023, doi: 10.1016/j.lwt.2023.114831.

[6] M. T. S. -Ballesta, C. M. -Anders, M. D. Álvarez, M. I. Escribano, C. Merodio, and I. Romero, "Are the blueberries we buy good quality? Comparative study of berries purchased from different outlets," *Foods*, vol. 12, no. 13, 2023, doi: 10.3390/foods12132621.

[7] A. A. -Cerezo, X. Yang, A. M. J. -Carvelo, M. Pellegrino, A. F. Savino, and P. Berzaghi, "Assessment of extra virgin olive oil quality by miniaturized near infrared instruments in a rapid and non-destructive procedure," *Food Chemistry*, vol. 430, 2024, doi: 10.1016/j.foodchem.2023.137043.

[8] C. S. P. Santos *et al.*, "Non-destructive measurement of the internal quality of citrus fruits using a portable NIR device," *Journal of AOAC International*, vol. 104, no. 1, pp. 61–67, 2021, doi: 10.1093/jaoacint/qsaa115.

[9] M. L. Montecchiarini *et al.*, "Proteomic and metabolomic approaches unveil relevant biochemical changes in carbohydrate and cell wall metabolisms of two blueberry (Vaccinium corymbosum) varieties with different quality attributes," *Plant Physiology and Biochemistry*, vol. 136, pp. 230–244, 2019, doi: 10.1016/j.plaphy.2018.12.019.

[10] P. Rungpichayapichet, B. Mahayothee, M. Nagle, P. Khuwijitjaru, and J. Müller, "Robust NIRS models for non-destructive prediction of postharvest fruit ripeness and quality in mango," *Postharvest Biology and Technology*, vol. 111, pp. 31–40, 2016, doi: 10.1016/j.postharvbio.2015.07.006.

[11] G. Ferrara, V. Marcotuli, A. Didonna, A. M. Stellacci, M. Palasciano, and A. Mazzeo, "Ripeness prediction in table grape cultivars by using a portable NIR device," *Horticulturae*, vol. 8, no. 7, 2022, doi: 10.3390/horticulturae8070613.

[12] S. Ditcharoen *et al.*, "Improving the non-destructive maturity classification model for durian fruit using near-infrared spectroscopy," *Artificial Intelligence in Agriculture*, vol. 7, pp. 35–43, 2023, doi: 10.1016/j.aiia.2023.02.002.

[13] A. Eldawlatly, H. Alshehri, A. Alqahtani, A. Ahmad, F. Al-Dammas, and A. Marzouk, "Appearance of population, intervention, comparison, and outcome as research question in the title of articles of three different anesthesia journals: A pilot study," *Saudi Journal of Anaesthesia*, vol. 12, no. 2, pp. 283-286, 2018, doi: 10.4103/sja.SJA_767_17.

[14] M. J. Page *et al.*, "Declaración PRISMA 2020: una guía actualizada para la publicación de revisiones sistemáticas," *Revista Española de Cardiología*, vol. 74, no. 9, pp. 790–799, Sep. 2021, doi: 10.1016/j.recesp.2021.06.016.

[15] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *Journal of Business Research*, vol. 133, pp. 285–296, 2021, doi: 10.1016/j.jbusres.2021.04.070.

[16] A. P.-Rodriguez, L. Waltman, and N. J. van Eck, "Constructing bibliometric networks: A comparison between full and fractional counting," *Journal of Informetrics*, vol. 10, no. 4, pp. 1178–1195, 2016, doi: 10.1016/j.joi.2016.10.006.

[17] M. Mancini *et al.*, "Application of near infrared spectroscopy for the rapid assessment of nutritional quality of different strawberry cultivars," *Foods*, vol. 12, no. 17, 2023, doi: 10.3390/foods12173253.

[18] W. Ye, W. Xu, T. Yan, J. Yan, P. Gao, and C. Zhang, "Application of near-infrared spectroscopy and hyperspectral imaging combined with machine learning algorithms for quality inspection of grape: a review," *Foods*, vol. 12, no. 1, 2023, doi: 10.3390/foods12010132.

[19] Y. Wu, X. Zhu, Q. Huang, Y. Zhang, J. Evans, and S. He, "Predicting the quality of tangerines using the GCNN-LSTM-AT network based on Vis–NIR spectroscopy," *Applied Sciences*, vol. 13, no. 14, 2023, doi: 10.3390/app13148221.

[20] B. Hasanzadeh, Y. A. -Gilandeh, A. S. -Nazarloo, E. D. L. C. -Gámez, J. L. H. -Hernández, and M. M. -Arroyo, “Non-destructive measurement of quality parameters of apple fruit by using visible/near-infrared spectroscopy and multivariate regression analysis,” *Sustainability*, vol. 14, no. 22, 2022, doi: 10.3390/su142214918.

[21] C. Guo, J. Zhang, W. Cai, and X. Shao, “Enhancing transferability of near-infrared spectral models for soluble solids content prediction across different fruits,” *Applied Sciences*, vol. 13, no. 9, 2023, doi: 10.3390/app13095417.

[22] M. Égei *et al.*, “Prediction of soluble solids and lycopene content of processing tomato cultivars by Vis-NIR spectroscopy,” *Frontiers in Nutrition*, vol. 9, 2022, doi: 10.3389/fnut.2022.845317.

[23] Y. Zhang, J. F. Nock, Y. Al Shoffe, and C. B. Watkins, “Non-destructive prediction of soluble solids and dry matter contents in eight apple cultivars using near-infrared spectroscopy,” *Postharvest Biology and Technology*, vol. 151, pp. 111–118, 2019, doi: 10.1016/j.postharvbio.2019.01.009.

[24] J. A. Martins *et al.*, “SpectraNet–53: A deep residual learning architecture for predicting soluble solids content with VIS–NIR spectroscopy,” *Computers and Electronics in Agriculture*, vol. 197, 2022, doi: 10.1016/j.compag.2022.106945.

[25] S. Sarkar, J. K. Basak, B. E. Moon, and H. T. Kim, “A comparative study of PLSR and SVM-R with various preprocessing techniques for the quantitative determination of soluble solids content of hardy kiwi fruit by a portable Vis/NIR spectrometer,” *Foods*, vol. 9, no. 8, 2020, doi: 10.3390/foods9081078.

[26] P. Pornchaloempong *et al.*, “Non-destructive quality evaluation of tropical fruit (mango and mangosteen) purée using near-infrared spectroscopy combined with partial least squares regression,” *Agriculture*, vol. 12, no. 12, 2022, doi: 10.3390/agriculture12122060.

[27] Y. Shao, G. Xuan, Z. Hu, Z. Gao, and L. Liu, “Determination of the bruise degree for cherry using Vis-NIR reflection spectroscopy coupled with multivariate analysis,” *PLoS ONE*, vol. 14, no. 9, 2019, doi: 10.1371/journal.pone.0222633.

[28] S. Xu, H. Lu, C. Ference, and Q. Zhang, “An accuracy improvement method based on multi-source information fusion and deep learning for tssc and water content nondestructive detection in ‘luogang’ orange,” *Electronics*, vol. 10, no. 1, pp. 1–12, 2021, doi: 10.3390/electronics10010080.

[29] A. Khumaidi, Y. A. Purwanto, H. Sukoco, and S. H. Wijaya, “Using fuzzy logic to increase accuracy in mango maturity index classification: approach for developing a portable near-infrared spectroscopy device,” *Sensors*, vol. 22, no. 24, 2022, doi: 10.3390/s22249704.

[30] E. Kalopesa, K. Karyotis, N. Tziolas, N. Tsakiridis, N. Samarinas, and G. Zalidis, “Estimation of sugar content in wine grapes via In Situ VNIR–SWIR point spectroscopy using explainable artificial intelligence techniques,” *Sensors*, vol. 23, no. 3, 2023, doi: 10.3390/s23031065.

[31] M. Shibang, “Nondestructive determination of kiwifruit SSC using visible/near-infrared spectroscopy with genetic algorithm,” *Journal of Engineering Science and Technology Review*, vol. 14, no. 1, pp. 100–106, 2021, doi: 10.25103/jestr.141.11.

[32] S. Xu, H. Lu, C. Ference, and Q. Zhang, “Visible/near infrared reflection spectrometer and electronic nose data fusion as an accuracy improvement method for portable total soluble solid content detection of orange,” *Applied Sciences*, vol. 9, no. 18, 2019, doi: 10.3390/app9183761.

[33] X. Yang *et al.*, “Determination of the soluble solids content in Korla fragrant pears based on visible and near-infrared spectroscopy combined with model analysis and variable selection,” *Frontiers in Plant Science*, vol. 13, 2022, doi: 10.3389/fpls.2022.938162.

[34] Z. Cai, S. Jiang, J. Dong, and S. Tang, “An artificial plant community algorithm for the accurate range-free positioning of wireless sensor networks,” *Sensors*, vol. 23, no. 5, 2023, doi: 10.3390/s23052804.

[35] A. Zeb *et al.*, “Towards sweetness classification of orange cultivars using short-wave NIR spectroscopy,” *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-022-27297-2.

[36] R. Pourdarbani, S. Sabzi, M. H. Rohban, G. G. -Mateos, J. Paliwal, and J. M. M. -Martínez, “Using metaheuristic algorithms to improve the estimation of acidity in Fuji apples using NIR spectroscopy,” *Ain Shams Engineering Journal*, vol. 13, no. 6, 2022, doi: 10.1016/j.asej.2022.101776.

[37] A. Puttipipatkajorn, A. Terdwongworakul, A. Puttipipatkajorn, S. Kulmutiwat, P. Sangwanangkul, and T. Cheepsomsong, “Indirect prediction of dry matter in durian pulp with combined features using miniature NIR spectrophotometer,” *IEEE Access*, vol. 11, pp. 84810–84821, 2023, doi: 10.1109/ACCESS.2023.3303020.

[38] P. Mishra and E. Woltering, “Semi-supervised robust models for predicting dry matter in mango fruit with near-infrared spectroscopy,” *Postharvest Biology and Technology*, vol. 200, 2023, doi: 10.1016/j.postharvbio.2023.112335.

[39] P. Escárate, G. Farias, P. Naranjo, and J. P. Zoffoli, “Estimation of soluble solids for stone fruit varieties based on near-infrared spectra using machine learning techniques,” *Sensors*, vol. 22, no. 16, 2022, doi: 10.3390/s22166081.

[40] J. A. Martins, D. Rodrigues, A. M. Cavaco, M. D. Antunes, and R. Guerra, “Estimation of soluble solids content and fruit temperature in ‘Rocha’ pear using Vis-NIR spectroscopy and the SpectraNet–32 deep learning architecture,” *Postharvest Biology and Technology*, vol. 199, 2023, doi: 10.1016/j.postharvbio.2023.112281.

[41] Y. Hao, X. Li, C. Zhang, and Z. Lei, “Research on construction method and validity mechanism of robust analysis model in snow peach quality detection based on visible-near infrared spectroscopy,” *Frontiers in Nutrition*, vol. 9, 2022, doi: 10.3389/fnut.2022.1042868.

[42] H. Galal, S. Elsayed, A. Allam, and M. Farouk, “Indirect quantitative analysis of biochemical parameters in banana using spectral reflectance indices combined with machine learning modeling,” *Horticulturae*, vol. 8, no. 5, 2022, doi: 10.3390/horticulturae8050438.

[43] S. S. R. M. Lazim, M. N. Nawi, S. K. Bejo, A. R. M. Shariff, and N. Abdullah, “Prediction and classification of soluble solid contents to determine the maturity level of watermelon using visible and shortwave near infrared spectroscopy,” *International Food Research Journal*, vol. 29, no. 6, pp. 1372–1379, 2022, doi: 10.47836/ifrj.29.6.13.

[44] F. P. Lamptey, E. Teye, E. E. Abano, and C. L. Y. Amuah, “Application of handheld NIR spectrometer for simultaneous identification and quantification of quality parameters in intact mango fruits,” *Smart Agricultural Technology*, vol. 6, 2023, doi: 10.1016/j.atech.2023.100357.

[45] M. Mancini *et al.*, “Prediction of soluble solids content by means of NIR spectroscopy and relation with botrytis cinerea tolerance in strawberry cultivars,” *Horticulturae*, vol. 9, no. 1, 2023, doi: 10.3390/horticulturae9010091.

[46] X. Jiang, M. Zhu, J. Yao, Y. Zhang, and Y. Liu, “Study on the effect of apple size difference on soluble solids content model based on near-infrared (NIR) spectroscopy,” *Journal of Spectroscopy*, vol. 2022, 2022, doi: 10.1155/2022/3740527.

[47] X. Sun, D. Deng, J. Liu, and S. Feng, “Model development and update of portable NIRS instrument for assessment of internal quality attributes of two navel orange varieties,” *Frontiers in Nutrition*, vol. 9, 2022, doi: 10.3389/fnut.2022.976178.

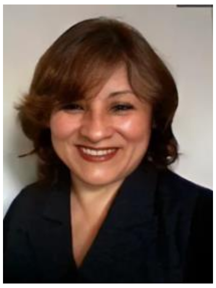
[48] P. Mishra, E. Woltering, B. Brouwer, and E. Hogeveen-van Echtelt, “Improving moisture and soluble solids content prediction in pear fruit using near-infrared spectroscopy with variable selection and model updating approach,” *Postharvest Biology and*

Technology, vol. 171, 2021, doi: 10.1016/j.postharvbio.2020.111348.
[49] M. Mancini *et al.*, "Application of the non-destructive NIR technique for the evaluation of strawberry fruits quality parameters," *Foods*, vol. 9, no. 4, 2020, doi: 10.3390/foods9040441.
[50] Z. Guo *et al.*, "Quantitative detection of apple watercore and soluble solids content by near infrared transmittance spectroscopy," *Journal of Food Engineering*, vol. 279, 2020, doi: 10.1016/j.jfoodeng.2020.109955.
[51] S. M. -S. Román, J. F. -Novales, C. C. -Tarancón, R. S. -Gómez, M. P. Diago, and T. G. -Cerdán, "Application of near-infrared spectroscopy for the estimation of volatile compounds in Tempranillo Blanco grape berries during ripening," *Journal of the Science of Food and Agriculture*, vol. 103, no. 13, pp. 6317–6329, 2023, doi: 10.1002/jsfa.12706.
[52] P. P. Subedi and K. B. Walsh, "Assessment of avocado fruit dry matter content using portable near infrared spectroscopy: Method and instrumentation optimisation," *Postharvest Biology and Technology*, vol. 161, 2020, doi: 10.1016/j.postharvbio.2019.111078.
[53] N. T. Anderson, K. B. Walsh, P. P. Subedi, and C. H. Hayes, "Achieving robustness across season, location and cultivar for a NIRS model for intact mango fruit dry matter content," *Postharvest Biology and Technology*, vol. 168, 2020, doi: 10.1016/j.postharvbio.2020.111202.
[54] T. Basile, A. D. Marsico, and R. Perniola, "Use of artificial neural networks and NIR spectroscopy for non-destructive grape texture prediction," *Foods*, vol. 11, no. 3, 2022, doi: 10.3390/foods11030281.
[55] C. L. Y. Amuah, E. Teye, F. P. Lamptey, K. Nyandey, J. O. -Ansah, and P. O. W. Adueming, "Feasibility study of the use of handheld NIR spectrometer for simultaneous authentication and quantification of quality parameters in intact pineapple fruits," *Journal of Spectroscopy*, vol. 2019, 2019, doi: 10.1155/2019/5975461.

BIOGRAPHIES OF AUTHORS

Liset Cayhualla Amaro is a student of Systems and Computer Engineering at the Technological University of Peru, Lima, Peru. She has a technical degree in Software Design and Development. As a student, he is expanding his knowledge in conducting research to implement new technologies, in order to have greater credibility in his professional career. She can be contacted at email: u21212941@utp.edu.pe.

Sebastian Rau Reyes is a student of the Technological University of Peru, Lima, Peru. He is currently studying Systems Engineering and Computer Science and also has a previous technical career in Software Design and Development. As a professional and student of his university, he seeks to gain new knowledge in the application of the latest information systems technologies related to artificial intelligence. He can be contacted at email: u21213403@utp.edu.pe.



María Acuña Meléndez is professor at the Faculty of Engineering of the Technological University of Peru, Lima, Peru. She has a Ph.D. in Systems Engineering. Her research areas are information systems and communications, as well as systems auditing and information security. He has participated in several research projects, as well as in thesis advising. She has registered several patents on software copyrights. Her research interests include information systems, data science, information security, data mining, artificial intelligence, knowledge management among other related lines of research. She can be contacted at email: c21584@utp.edu.pe.

Christian Ovalle is an associate professor at the Faculty of Engineering of the Technological University of Peru, Lima, Peru. He has a Ph.D. in Systems Engineering with specialization in artificial intelligence. His research areas are process mining, business data analysis and pattern recognition. He is CEO of the consulting firm 7D dedicated to Intelligent Solutions Research. He has participated in different research projects, receiving awards from the Ministry of Defense of Peru and the Army of the Armed Forces for the best general researcher, which is a technology-based company and its innovative products received national and international recognition. He has filed several patents and industrial designs on his innovative ideas. His research interests include data mining, artificial intelligence, image/signal processing, bibliometrics, and pattern recognition. He can be contacted at email: dovalle@utp.edu.pe.