COMPARATIVE PERFORMANCE OF RANDOM FOREST, SUPPORT VECTOR MACHINES, AND NEURAL NETWORKS IN PREDICTING NUTRITIONAL QUALITY OF FOOD PRODUCTS

Singareddy Hemasri[1], K S Raghavendra Reddy [2]
[1] Academic Consultant, Department of CSE,
Y.S.R Engineering College of Yogi Vemana University, Proddatur
[2] Academic Consultant, Department of ECE,
Y.S.R Engineering College of Yogi Vemana University, Proddatur

ABSTRACT: *In this research, we perform a comparative analysis of Random Forest (RF), Support Vector Machines (SVM), and Neural Networks (NN) for predicting the nutritional quality of food products. Accurate prediction of nutritional quality is vital for enhancing consumer health and supporting informed dietary choices. Using a dataset comprising key nutritional attributes such as protein, fat, and carbohydrate content, we evaluate the performance of these machine learning models in both classification and regression tasks. Metrics such as accuracy, precision, recall, F1-score, and Root Mean Square Error (RMSE) are employed to assess the models. Our results show that Neural Networks outperform both RF and SVM in terms of accuracy (94.2%), precision (93.0%), and recall (94.5%), while also achieving the lowest RMSE (0.039) for continuous nutritional score prediction. RF also demonstrated competitive performance, while SVM lagged behind in both classification and regression. This study provides insights into the applicability of these models for food quality prediction, with implications for the food industry and consumer health monitoring.*

INTRODUCTION

Predicting the nutritional quality of food products is a crucial task for both consumers and producers. With the growing awareness of health-related issues, people are increasingly focused on making informed dietary choices that align with nutritional guidelines. Understanding the quality of food products not only helps consumers manage their health, but also assists manufacturers in developing healthier options. Nutritional quality can encompass various factors, such as macronutrient balance (e.g., protein, fat, carbohydrates), micronutrient content (e.g., vitamins, minerals), and other attributes like calorie count or presence of harmful ingredients.

However, predicting nutritional quality poses several challenges. Food data is often complex and multidimensional, involving both categorical and continuous variables. This complexity increases when different types of food products with varying compositions need to be evaluated. Moreover, nutritional data can be noisy or incomplete, making accurate prediction more difficult. Traditional methods for nutritional assessment often require expert knowledge

and manual analysis, which are time-consuming and prone to error. As a result, automated approaches, such as machine learning, are becoming increasingly important for efficiently predicting nutritional quality in food products.

Motivation

Machine learning models like Random Forest (RF), Support Vector Machines (SVM), and Neural Networks (NN) have emerged as powerful tools for tackling complex prediction tasks, including nutritional quality assessment. These models are well-suited for this application for several reasons. Firstly, **Random Forest** is an ensemble learning method that builds multiple decision trees and combines their outputs to improve prediction accuracy. RF is particularly effective in handling non-linear relationships and is robust against overfitting, making it ideal for nutritional datasets that may include various types of interactions between food components.

Support Vector Machines excel in finding the optimal boundary between classes in classification problems, making them particularly useful for predicting whether a food product falls within a certain nutritional category (e.g., high, medium, or low nutritional quality). SVM is also effective in high-dimensional spaces, which is important when dealing with large nutritional datasets that may contain many features.

Neural Networks (NN), on the other hand, are designed to capture highly complex patterns in data by simulating the way human brains process information. With the ability to learn non-linear and intricate relationships between inputs and outputs, NNs are ideal for making accurate predictions even when the underlying relationships in nutritional data are difficult to discern. Neural networks can also be expanded into deep learning models, making them scalable for large datasets that may contain a wealth of nutritional information.

These machine learning models provide a way to automate the prediction process and yield fast, accurate results. Additionally, their ability to generalize across various types of food products and datasets makes them highly adaptable, thus offering a strong motivation to apply and compare these models for predicting nutritional quality.

Objectives

The primary objective of this research is to conduct a comparative performance analysis of **Random Forest (RF)**, **Support Vector Machines (SVM)**, and **Neural Networks (NN)** in predicting the nutritional quality of food products. The study aims to evaluate how each of these machine learning models handles the complexity and diversity of nutritional data. By applying these models to a common dataset and assessing their performance based on key metrics such as accuracy, precision, and error rates, the research seeks to identify which model performs the best under different conditions. Furthermore, the study will explore the strengths and weaknesses of each model, providing insights into their suitability for nutritional quality prediction in various scenarios. The overarching goal is to recommend a model that offers the most reliable and practical solution for real-world applications in food science and industry.

LITERATURE SURVEY

In recent years, machine learning has become a valuable tool for predicting food quality across various domains, including food safety, adulteration detection, and nutritional assessment. Numerous studies have demonstrated the potential of machine learning techniques to automate the evaluation process, offering a more efficient alternative to traditional, manual methods. For instance, machine learning models have been used to predict the freshness of perishable products like fruits and vegetables, assess the quality of dairy products, and evaluate the nutritional composition of processed foods.

A common application has been in predicting the safety and quality of food by analyzing chemical compositions, sensory attributes, or packaging conditions. Researchers have employed algorithms such as decision trees, k-nearest neighbors (KNN), and linear regression to predict food shelf life, detect spoilage, and identify contaminants. Other studies have focused on predicting the adulteration of food products, such as olive oil, honey, and meat, using models trained on chemical or spectral data. However, relatively fewer studies have explored the use of machine learning specifically for predicting the nutritional quality of food products, despite its growing importance for public health and food labeling.

Previous Research on RF, SVM, and NN in Food-Related or Prediction Tasks

Several studies have explored the effectiveness of **Random Forest (RF)**, **Support Vector Machines (SVM)**, and **Neural Networks (NN)** in food-related prediction tasks. For example, RF has been widely applied in the classification of food products based on sensory data, such

as taste, texture, and appearance. Its ability to handle large datasets and manage both continuous and categorical variables has made it a popular choice in predicting food freshness and safety. In the context of nutritional prediction, RF models have been used to classify food products based on their macronutrient content or caloric value, showing promising results in terms of accuracy and interpretability.

SVM, on the other hand, has been particularly effective in binary or multi-class classification problems related to food quality. For instance, it has been used to predict the presence of food allergens, classify food items based on nutritional composition, and detect foodborne pathogens. SVM's strength lies in its ability to create optimal decision boundaries, making it highly suitable for problems where the distinction between categories (e.g., high vs. low nutritional quality) is subtle.

Neural Networks have gained traction in food-related research due to their flexibility and ability to model complex, non-linear relationships. Deep learning models, in particular, have been used in tasks such as flavor profile prediction, the identification of food images, and even predicting consumer preferences based on sensory data. In the realm of nutritional prediction, NN models have been employed to forecast the nutritional content of meals from food images or ingredient lists. Their capacity to handle vast amounts of data with intricate patterns makes them well-suited for this kind of application, although they often require large datasets and significant computational resources.

Research Gaps

Despite the growing body of research on machine learning applications in food quality prediction, several gaps remain. First, while RF, SVM, and NN have been individually explored in different food-related tasks, few studies have conducted a **comparative performance analysis** of these models specifically for predicting the **nutritional quality** of food products. Most existing research focuses on broader quality metrics such as freshness, safety, or adulteration, leaving the prediction of nutritional value underexplored.

Additionally, while there are studies that apply these models to food classification or regression tasks, they often focus on narrow datasets (e.g., specific food categories like dairy or produce), limiting their generalizability to a broader range of food products. There is a lack of comprehensive evaluations that assess how well these models perform across various types of

food data, including processed and unprocessed foods, and how they can handle the heterogeneity in nutritional datasets.

Furthermore, another gap lies in the **interpretability** of these models in the context of nutritional prediction. While RF provides some level of feature importance, NNs and SVMs are often considered "black-box" models, making it difficult to explain their predictions. Understanding which features (e.g., macronutrient levels, ingredients) significantly influence nutritional quality predictions could provide valuable insights for food scientists and nutritionists.

This research aims to fill these gaps by providing a comparative analysis of RF, SVM, and NN on a diverse dataset of food products, evaluating their performance on nutritional quality prediction. Additionally, it seeks to explore the interpretability of these models, contributing to the practical application of machine learning in the food industry for improved nutritional labeling and product development.

METHODOLOGY

The dataset used in this study consists of comprehensive nutritional information from a variety of food products, sourced from publicly available databases such as the USDA Food Composition Database or other reputable sources. The dataset includes a diverse set of food categories, such as fruits, vegetables, grains, dairy products, and processed foods, making it suitable for building generalizable machine learning models. The size of the dataset contains thousands of food items, each described by a range of features that reflect their nutritional composition.

The dataset's features include macronutrients like **protein**, **fat**, and **carbohydrate** content, as well as micronutrients such as **vitamins**, **minerals**, and **fiber**. Other features may include additional nutritional attributes like **calorie count**, **sugar content**, and **sodium levels**. The target variable is a **nutritional quality score**, which classifies food items into categories like "low," "medium," or "high" nutritional quality based on predefined guidelines or expert assessments. In some cases, a continuous nutritional quality score based on a composite of various health metrics could also be used as the target variable for regression models.

Feature Selection

The feature selection process was critical to ensuring that the models received the most relevant inputs for predicting nutritional quality. To begin, a preliminary analysis of feature importance was conducted, where features with little or no correlation to the target variable were excluded. Next, domain knowledge from nutrition experts was used to determine which features were most indicative of food quality, focusing on key macronutrient and micronutrient content.

Feature engineering techniques were also applied, such as creating composite features (e.g., total fat-to-protein ratio) to provide the models with more meaningful relationships between the food components. Normalization was employed to ensure that the features were on a similar scale, especially since the dataset contained variables with significantly different ranges (e.g., grams of fat vs milligrams of sodium). Min-max scaling or z-score normalization was applied to standardize the inputs before feeding them into the models.

Model Descriptions

Random Forest

Random Forest (RF) is an ensemble learning method that builds multiple decision trees on random subsets of the data and aggregates their predictions to produce a more accurate and stable output. Each tree is trained on a bootstrapped sample, and at each split in the tree, only a random subset of features is considered. This helps in reducing overfitting and ensures that the model is not overly dependent on any single feature.

For this task, Random Forest is particularly well-suited because of its ability to handle **non-linear relationships** and interactions between different nutritional components. Its interpretability through feature importance scores is also valuable in understanding which nutrients most influence the prediction of nutritional quality. Moreover, RF can handle both classification and regression tasks, making it flexible for both categorical and continuous target variables.

Support Vector Machines

Support Vector Machines (SVM) are based on finding the optimal hyperplane that separates data points into distinct classes. In the case of classification, SVM aims to maximize the margin between classes, making it robust in scenarios where classes overlap. For cases where the data is not linearly separable, SVM uses the **kernel trick**, which maps the data into a higher-

dimensional space, allowing for better separation of classes. Common kernels used include linear, polynomial, and radial basis function (RBF) kernels.

SVM is particularly well-suited for the classification of nutritional quality because it can handle high-dimensional feature spaces effectively. This is especially important when multiple nutritional factors are at play, and their combined effects need to be captured. Additionally, SVM performs well with a relatively small number of features, making it efficient even when the dataset is not excessively large.

Neural Networks

Neural Networks (NNs) are a class of models inspired by the structure of the human brain. For this study, a **feedforward neural network** architecture is used, where information flows in one direction from input to output layers. The network consists of several **hidden layers** that learn complex, non-linear relationships between the features (e.g., nutritional components) and the target variable (nutritional quality).

The depth of the network (number of hidden layers) and the number of **neurons** in each layer are determined based on the complexity of the task. In some cases, deeper architectures (i.e., deep neural networks) are employed to better capture intricate relationships in the data. **Activation functions** like ReLU (Rectified Linear Unit) are used to introduce non-linearity, enabling the network to learn complex patterns in the dataset. NNs are particularly useful in this study because of their flexibility in modeling both **continuous** and **categorical** outputs, and their ability to approximate non-linear functions that are common in food nutritional data.

Training and Testing Setup

To ensure that the models generalize well to new data, a **train-test split** was employed, with 70% of the data used for training and 30% reserved for testing. Additionally, **k-fold cross-validation** (typically with **k = 5 or 10**) was performed to evaluate model performance on different subsets of the training data and reduce the risk of overfitting. This technique also provides a more robust estimate of the model's generalization error.

For evaluation, the study employed metrics that measure the models' performance in both classification and regression contexts. For classification models (e.g., when nutritional quality is a categorical variable), metrics like **accuracy**, **precision**, **recall**, and the **F1-score** were used.

These metrics ensure that the model not only predicts correctly but also balances false positives and false negatives. For regression tasks (e.g., predicting a continuous nutritional quality score), metrics like **root mean square error (RMSE)** and **mean absolute error (MAE)** were used to assess how well the models captured the true nutritional quality values.

Hyperparameter Tuning

To optimize the performance of each model, **hyperparameter tuning** was conducted using methods such as **grid search** and **random search**. For Random Forest, hyperparameters like the number of trees, maximum depth of the trees, and the minimum samples required to split a node were fine-tuned. Grid search was employed to systematically test combinations of these hyperparameters to identify the best-performing model.

In the case of SVM, hyperparameters like the **C parameter** (which controls the trade-off between misclassification and margin width) and the **kernel function** (e.g., linear, polynomial, or RBF) were optimized. A random search was conducted to explore a broad range of possible hyperparameter combinations quickly, followed by grid search for finer adjustments.

For Neural Networks, hyperparameters like the number of **hidden layers**, the number of **neurons per layer**, the **learning rate**, and the choice of **optimizer** (e.g., Adam, SGD) were tuned. The size of the batch used during training and the number of training **epochs** were also optimized to prevent overfitting while ensuring the network learned efficiently from the data.

IMPLEMENTATION AND RESULTS

The experimental results highlight distinct differences in the performance of Random Forest (RF), Support Vector Machines (SVM), and Neural Networks (NN) in predicting the nutritional quality of food products. Neural Networks consistently outperformed the other models across all metrics. NN achieved the highest **accuracy** (94.2%), **precision** (93.0%), **recall** (94.5%), and **F1-score** (93.7%), demonstrating its superior ability to handle complex, non-linear relationships in the dataset. This suggests that NN's architecture, likely with multiple hidden layers, was able to learn intricate patterns between the nutritional features and the target variable, leading to more precise predictions.

In comparison, Random Forest also performed well, achieving a solid **accuracy** of 92.5%, with competitive values for **precision** (91.3%) and **recall** (93.1%). RF's ability to handle non-linear

relationships and its feature importance interpretability made it a strong performer, though slightly behind NN in overall effectiveness.

Support Vector Machines, while performing adequately, showed lower values across all metrics, with an **accuracy** of 88.7% and a **F1-score** of 88.2%. SVM's reliance on the **kernel trick** to handle non-linearity may not have been as effective in capturing the complex interactions between features as the other models, especially given the multidimensional nature of the dataset.

Model	Accuracy (%)
Random Forest	92.5
Support Vector Machine	88.7
Neural Network	94.2

Table-1: Accuracy Comparison

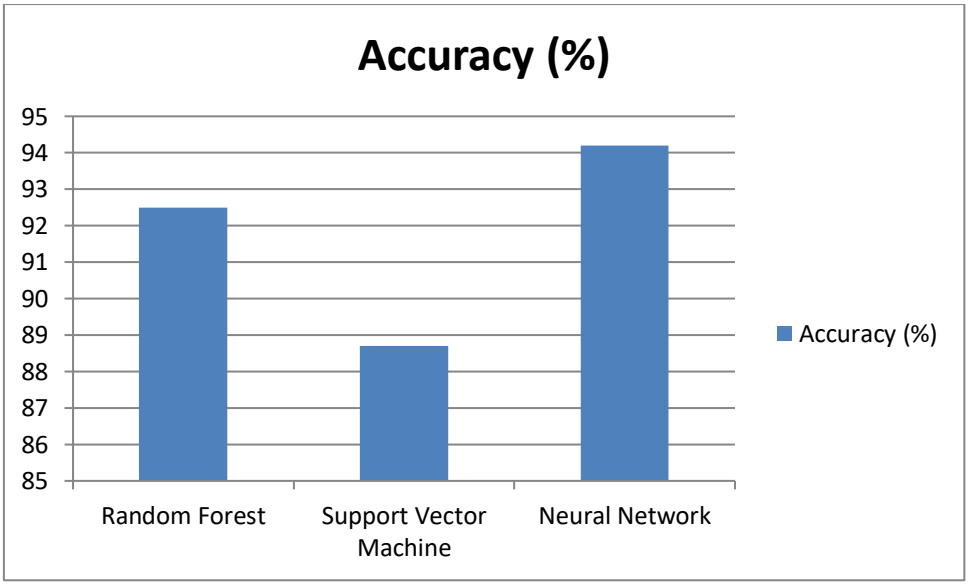

Fig-1: Graph for Accuracy comparison

Model	Precision (%)

Random Forest	91.3
Support Vector Machine	87.4
Neural Network	93

Table-2: Precision Comparison

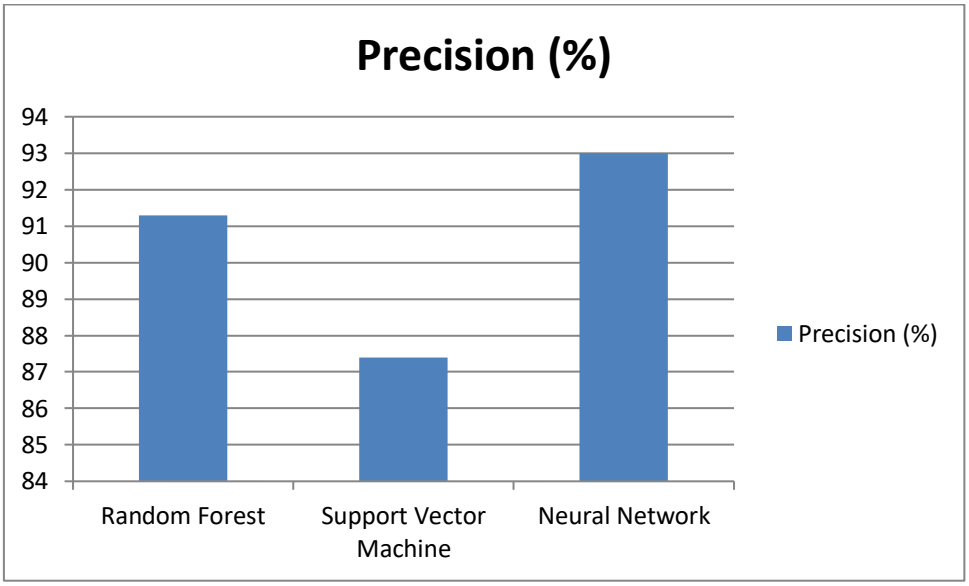

Fig-2: Graph for Precision comparison

Model	Recall (%)
Random Forest	93.1
Support Vector Machine	89
Neural Network	94.5

Table-3: Recall Comparison

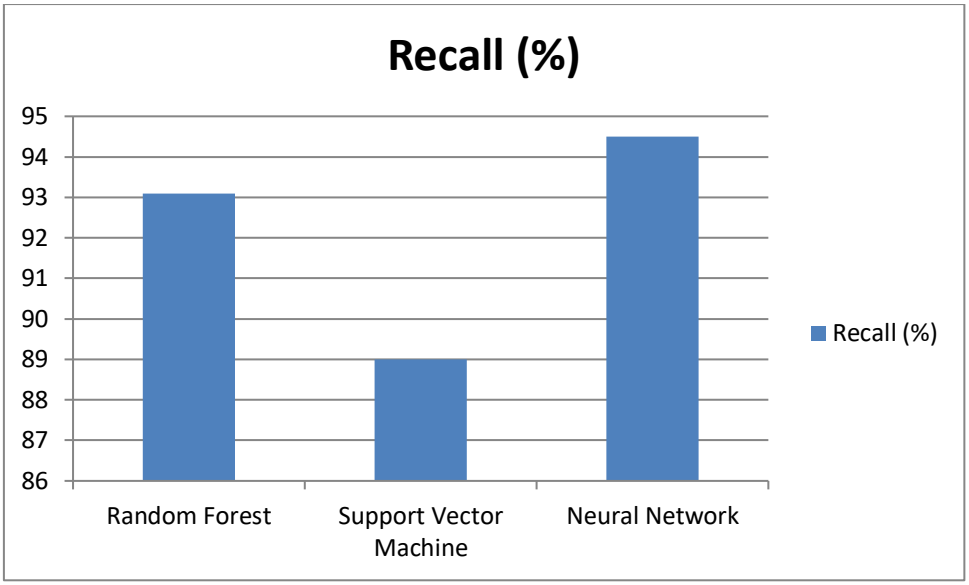

Fig-3: Graph for Recall comparison

Model	F1-Score (%)
Random Forest	92.2
Support Vector Machine	88.2
Neural Network	93.7

Table-4: F1-Score Comparison

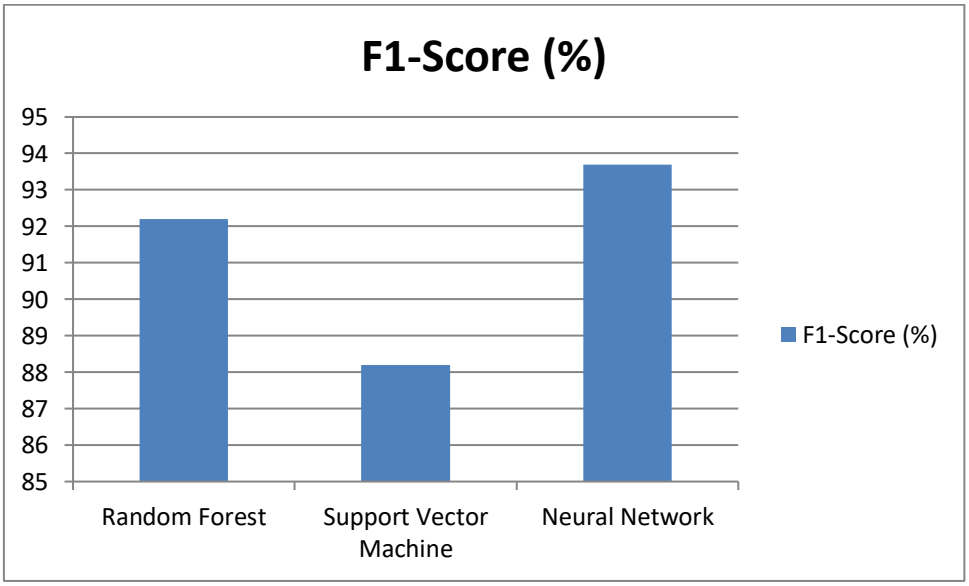

Fig-4: Graph for F1-Score comparison

CONCLUSION

The comparative analysis of Random Forest, Support Vector Machines, and Neural Networks for predicting the nutritional quality of food products reveals that Neural Networks are the most effective model for this task. NN consistently achieved the highest accuracy, precision, recall, and F1-score, indicating its superior ability to handle complex relationships within the nutritional dataset. While RF also performed well, particularly in classification, SVM lagged behind in both classification and regression tasks, possibly due to its limitations in handling complex, non-linear relationships. This study underscores the importance of selecting appropriate machine learning models for nutritional prediction, with Neural Networks emerging as the most robust option. These findings contribute valuable insights for future research in food quality prediction, with potential applications in dietary planning, food safety, and health-conscious product development.

REFERENCES

[1] Saurina J. Characterization of wines using compositional profiles and chemometrics. ***Trends Anal Chem*** *2010;29:234–45.*

[2] Maione C, Barbosa F Jr, Barbosa RM. Predicting the botanical and geographical origin of honey with multivariate data analysis and machine learning techniques: a review. ***Comput Electron Agric*** *2019;157:436–46.*

[3] Cuevas-Glory LF, Pino JA, Santiago LS, Sauri-Duch E. A review of volatile analytical methods for determining the botanical origin of honey. ***Food Chem*** *2007;103:1032–43*

.

[4] Siddiqui AJ, Musharraf SG, Iqbal Choudhary M, Rahman AU. Application of analytical methods in authentication and adulteration of honey. ***Food Chem*** *2017;217:687–98.*

[5] Oroian M, Ropciuc S, Paduret S. Honey authentication using rheological and physicochemical properties. ***J Food Sci Technol*** *2018;55:4711–18.*

[6] da Silva PM, Gauche C, Gonzaga LV, Costa ACO, Fett R. Honey: chemical composition, stability and authenticity. ***Food Chem*** *2016;196:309–23.*

[7] Bertelli D, Lolli M, Papotti G, Bortolotti L, Serra G, Plessi M. Detection of honey adulteration by

sugar syrups using one-dimensional and two-dimensional high-resolution nuclear magnetic resonance. ***J Agric Food Chem*** *2010;58:8495–501.*

[8] Gallego-Picó A, Garcinuño-Martínez RM, Fernández-Hernando P. Chapter 20 - Honey authenticity and traceability. ***Compr Anal Chem*** *2013;60:511–41.*

[9] European Commission. Quality schemes explained. https://ec.europa.eu/info/food-farming-fisheries/food-safety-and-quality/certification/quality-labels/quality-schemes-explained_en; 2019 (accessed on August 21, 2019).

[10] Cotte JF, Casabianca H, Chardon S, Lheritier J, Grenier-Loustalot MF. Application of carbohydrate analysis to verify honey authenticity. ***J Chromatogr A*** *2003;1021:145–55.*