AI-supported Identification of Vegetable Fats and Oils

Christian Gertz

Maxfry GmbH, Hagen, Germany

The authentication of edible oils has become increasingly important for ensuring product quality, safety, and compliance with regulatory standards. Adulteration of high-priced oils with cheaper alternatives or inferior oils, mislabeling the oil source or grade, and misrepresenting the origin of the oil are some common problems with authenticity. The authenticity testing of edible oils is usually carried out by comparing fatty acid spectra and other analytical characteristics with reference values in standards such as Codex Alimentarius.

In the new approach, the oleic acid content, iodine value, and saponification value (SPV) were determined in 2,600 samples of a total of 29 different vegetable fats and oils. After adding a fourth component *k* (*k* = C18:1 × SPV/IV), which is calculated from IV, SPV, and C18:1, all 29 oils can be differentiated and totally identified using *k* despite different proportions. The following oils were analyzed: almond, argan, avocado, coconut, corn, grapeseed, hazelnut, hemp, high oleic sunflower, linseed, lupin seed, macadamia, olive, pomace, palm, palmoleine, palmstearin, peanut, pumpkin seed, rape seed, HO-rape Seed, sunflower oil, safflor, HO-safflor, sesame, soybean, sunflower, walnut, wheat germ.

A database of these oils with 6 central objects of each oil type (calculated by k-means clustering) is the basis for the AI-supported identification. The statistical evaluation is carried out using Random Forest. Random Forest is a commonly used machine learning algorithm that combines the output of multiple decision trees into a single result. This method can also be used to determine the type and proportion of the foreign oil such as hazelnut, high oleic sunflower, pomace, or rapeseed oil in olive oil. Hereby, different proportions (10–50%) of a foreign oils are mixed with olive oil as references, and C18:0, C18:2, IV, and SPV have to be determined.

In the same way, an AI method was developed for olive oil analyzing the fatty acids C16:0, C16:1, C18:1(9c), C18:2 and the triacylglycerols POP, POO, PLP, OOO, OLO, and OLL, with more than 20,000 data records to differentiate and identify 29 different production countries, the corresponding geographical regions, and the processed varieties.