Autenticazione del pepe nero mediante l’ utilizzo di spettroscopia NIR, classificatore LASSO ed elaborazione di un’applicazione web per un facile screening su larga scala.

**Marco Bragolusi1**\***, Andrea Massaro1, Alessandra Tata1, Stephane Lefevre2, Jean-Louis Lafeuille3, Aline Fregiere Salomon2, Carmela Zacometti1, Giuseppe Sammarco4, Ingrid Fiordaliso Candalino5, Michele Suman4, Roberto Piro1.**

1Istituto Zooprofilattico Sperimentale delle Venezie, Laboratorio di Chimica Sperimentale Viale Fiume, 78, Vicenza Italy

2 Food Integrity Laboratory, Global Quality and Food Safety Center of Excellence, McCormick & Co., Inc., 999 avenue des Marchés, 84200 Carpentras, France

3 Global Quality and Food Safety Center of Excellence, McCormick & Co., Inc., 999 avenue des Marchés, 84200 Carpentras, France

4 Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Via Mantova, 166, 43122, Parma, Italy

5 Global Quality and Food Safety Center of Excellence, McCormick & Co., Inc., Viale Iotti Nilde, 50038 San Piero (FI), Italy

\*Corresponding author

Il pepe nero è una spezia soggetta ad adulterazioni per scopi economici (EMA). Questo contributo descrive un metodo non-targeted per l'autenticazione del pepe nero mediante spettroscopia del vicino infrarosso (NIR) unita ad un modello statistico basato su un least absolute shrinkage and selection operator (LASSO). Inizialmente sono stati analizzati 116 campioni tra cui pepe nero autentico (proveniente da diverse regioni asiatiche, africane e sudamericane e diverse stagioni di raccolta) e pepe addizionato con diversi adulteranti, inclusi scarti della pianta e 16 diversi materiali esogeni (semi di papaia, coriandolo, gesso, anice, nocciolo d'oliva, talco, riso, mais, senape, lenticchie, fagioli, peperoncino, aglio). La percentuale di adulterazione variava tra l'1,5% e il 35%.

Gli spettri sono stati suddivisi in training e test set e normalizzati mediante correzione moltiplicativa della dispersione (MSC). Mentre il test set è stato accantonato per successivo test del modello, il training set è stato sottoposto a LASSO con l'obiettivo di classificare i campioni autentici, adulterati con materiale esogeno ed prodotti di scarto (materiale endogeno). Il modello LASSO ha permesso la selezione e il restringimento delle variabili significative. Il modello risultante è stato validato mediante cross-validation ottenendo un'elevata accuratezza. Il modello è stato testato sul test set ottenendo un'accuratezza complessiva del 94% e tassi di sensibilità e specificità molto elevati. Il modello è stato poi validato con 34 campioni indipendenti e ne ha classificato correttamente 33/34. Il modello è stato testato con un proficiency test. Sono in corso ulteriori verifiche interlaboratorio.

E’ stata realizzata un'applicazione di web Shiny per la classificazione direttamente dai dati grezzi. L’operatore NIR carica i dati grezzi nell'app che esegue la normalizzazione e l'interrogazione del modello LASSO. L'applicazione online facilita la fase di classificazione e restituisce una chiara visualizzazione dei risultati. Questa è la prima applicazione di LASSO ai dati di spettroscopia NIR.

**Keywords:** Adulterazione, machine learning, frode.

Combination of NIR spectroscopy and LASSO modelling for black pepper authentication: development of the method, exploration of validation strategies and build-up of a user-friendly online application for large-scale screening.

**Marco Bragolusi1**\***, Andrea Massaro1, Alessandra Tata1, Stephane Lefevre2, Jean-Louis Lafeuille3, Aline Fregiere Salomon2, Carmela Zacometti1, Giuseppe Sammarco4, Ingrid Fiordaliso Candalino5, Michele Suman4, Roberto Piro1.**

1Istituto Zooprofilattico Sperimentale delle Venezie, Laboratorio di Chimica Sperimentale Viale Fiume, 78, Vicenza Italy

2 Food Integrity Laboratory, Global Quality and Food Safety Center of Excellence, McCormick & Co., Inc., 999 avenue des Marchés, 84200 Carpentras, France

3 Global Quality and Food Safety Center of Excellence, McCormick & Co., Inc., 999 avenue des Marchés, 84200 Carpentras, France

4 Advanced Laboratory Research, Barilla G. e R. Fratelli S.p.A., Via Mantova, 166, 43122, Parma, Italy

5 Global Quality and Food Safety Center of Excellence, McCormick & Co., Inc., Viale Iotti Nilde, 50038 San Piero (FI), Italy

\*Corresponding author

Black pepper is a valuable commodity susceptible to economically motivated adulterations. This contribution describes a non-targeted method for authentication of black pepper by near infrared spectroscopy (NIR) coupled to least absolute shrinkage and selection operator (LASSO). An amount of 116 samples were collected and analyzed. The dataset included authentic black pepper samples (from different Asian and South American regions and four harvesting seasons) and samples spiked with different adulterants, including both endogenous sub-products (pinhead and spent) and exogenous materials (papaya seeds, coriander, plaster, green anise, olive kernel, talc, rice flour, black mustard, green lentil, red beans, chili, garlic and corn flour). The percentage of adulteration ranged between 1.5% and 35%.

The spectra were split in training and test set and then normalized by multiplicative scatter correction (MSC). While the test set was withheld for further testing of the model, the training set was submitted to LASSO with the aim to classify the samples as authentic, exogenously-adulterated and endogenously-adulterated group. LASSO model is used for selection and shrinkage of parameters. The resulting model was cross-validated achieving high overall accuracy. The model was tested on the test set achieving an overall accuracy of 94% and very high sensitivity and specificity rates. The model was then validated with a batch of 34 independent samples and it classified correctly 33/34 samples. The model was also challenged with samples from an inter-laboratory proficiency test. Multiple laboratory verifications of the proposed methodology are ongoing.

Finally, a Shiny web application was set up for direct statistical analysis of the raw data. The NIR user simply uploads the raw data to the app and MSC normalization and interrogation of the LASSO model is performed. The online application facilitates the model testing and enables clear results visualization. This is the first application of LASSO to spectroscopy data.

**Keywords:** Adulteration, machine learning, fraud