İstanbul Journal of Pharmacy
Istanbul J Pharm, 53 (3): 320–328
DOI: 10.26650/IstanbulJPharm.2023.1181298

Original Article

A chemometrics-based approach for the determination of thymoquinone from *Nigella sativa* L. (Black Cumin) seeds of different geographical regions using the HPLC technique

Selin Isık[1], Abdullahi Garba Usman[1,2], Sani Isah Abba[3]

[1]Near East University, Faculty of Pharmacy,Department of Analytical Chemistry, Nicosia, Cyprus
[2]Near East University, Operational research Centre in Healthcare, Nicosia, Cyprus
[3]King Fahd University of Petroleum and Minerals Interdisciplinary Research Center for Membrane and Water Security, Dhahran, Saudi Arabia

ABSTRACT

Background and Aims: In this study, thymoquinone (TQ) from black cumin will be quantified from several geographical regions, including India, Syria, Saudi Arabia, Iraq, and Turkey. Additionally, to forecast the chromatographic behavior of the analyte in artificial intelligence (AI)-based models, the study used both ensemble machine learning methodologies and chemometrics-based approaches.

Methods: An Agilent Technologies (1200 series, USA) instrument that includes an autosampler, a binary pump, a diode array detector (DAD), and a vacuum degasser was used for the HPLC analysis. Using five different single models—principal component regression (PCR), least square-support vector machine (LSSVM), least square boost (LSQ-BOOST), adaptive neuro-fuzzy inference system (ANFIS), and step-wise linear regression—the HPLC-DAD technique was used to simulate the qualitative and quantitative properties of TQ (SWLR).

Results: The collected results demonstrated that samples from India and Iraq have the highest concentration of TQ. TQ was present in all samples, but in varying amounts; the amounts of TQ in the samples from Iraq, India, Saudi Arabia, Syria, and Turkey, respectively, were 0.031, 0.030, 0.022, 0.005, and 0.001%. According to a comparison of their performances, the four ensemble machine learning techniques can reproduce the chromatographic properties of TQ, PA, and tR with minimum and maximum NSE-values of 0.842 and 0.999 in the training phase and 0.918 and 0.999 in the testing phase, respectively.

Conclusion: The TQ content of each sample of black cumin, which was collected from various geographical locations, was determined quantitatively. The quantity of thymoquinone fluctuates depending on geographic variances, according to HPLC data. Five different AI-based models, including SWLR, PCR, LSSVM, ANFIS, and LSQ-Boost, were used to simulate the chromatographic behavior of TQ information of retention duration and peak area using various independent factors. Additionally, SAE, WAE, NNE, and ANFIS-E are informed by the application of ensemble machine learning to enhance the performance of AI-based models. Comparing the approaches to the individual models, they both demonstrated lower error values in terms of RMSE and MSE.

Keywords: Chemometrics, HPLC, Thymoquinone, Black cumin, geographical regions

INTRODUCTION

Black cumin's major bioactive compound, thymoquinone (TQ), has a wide range of biological effects, including anti-cancer, antioxidant, anti-inflammatory, and hepatoprotective qualities. These effects are hypothesized and supported by evidence from science that illustrates the molecular mechanism of the analyte (Rezai, Işk, Kartal, & Aslan Erdem, 2018). Finding TQ in Black cumin and other medicinal plants is regarded as a challenging and time-consuming task due to the huge quantity of bioactive components that are present in a plant. Depending on the region, the type of plant material, and the objective of the investigation, many methods, such as chemical, microbiological, biological, and chromatographic ones, can be utilized. Although microbiological techniques are the most used, their lack of specificity is making them obsolete (Chen et al., 2019).

Chromatographic techniques like High-performance liquid chromatography (HPLC) are capable of separating various analytes in a plant with a high degree of specificity and sensitivity,

Corresponding Author: Selin Işık **E-mail:** selin_isikk@hotmail.com
Submitted: 28.09.2022 • **Revision Requested:** 21.10.2022 • **Last Revision Received:** 14.09.2023 • **Accepted:** 22.09.2023

making them useful for isolating TQ from Black cumin. For appropriate resolution, it's imperative to optimize several chromatographic factors, including the mobile phase concentration, pH of the mobile phase, column temperature, mobile phase flow rate, and column type. For example, achieving appropriate resolution requires optimizing several chromatographic factors, including the mobile phase concentration, mobile phase pH, column temperature, mobile phase flow rate, and column type. The experimental approach is linked to artificial intelligence (AI) in a comprehensive optimization strategy that has recently been put forth (Marrero-Ponce, Barigye, Jorge-Rodrguez, & Tran-Thi-Thu, 2018). Following a proper experimental procedure, chemometrics approaches are used to forecast the outcomes of the experiment using various input and output characteristics, and the coupling is deemed successful. The application of chemometrics and AI-based models in the development of HPLC methods is widely established in the technical literature. For instance, Tham and Agatonovic-Kustrin combined a genetic algorithm with an artificial neural network (ANN) to forecast the retention behavior of phenylthiocarbamyl amino acid derivatives (Tham & Agatonovic-Kustrin, 2002). According to Vasiljevi et al., HPLC method development has been used to optimize ANN for simulating the retention behavior of various analytes. According to their findings (Vasiljevi, Onjia, Okea, & Lauevi, 2004), ANN can be used to estimate retention time in an isocratic elution system for separating mixtures of complex compounds with significant changes in log Kow and pKa values. Support vector machines (SVM) and artificial neural networks (ANN) are used by Golmohammad et al. to analyze the retention times of different peptides using an HPLC system. The top models were characterized using a combination of Genetic Algorithm and Partial Least Squares (GA-PLS) optimization techniques. The findings demonstrated that SVM performed better than all other models in all of the data sets utilized in the study and increased the accuracy of prediction performance (Golmohammadi, Dashtbozorgi, & Vander Heyden, 2015). The application of the heuristic method (HM) and SVM in the creation and evaluation of linear and non-linear models for simulating retention time and molecular predictors of various volatile organic molecules was also covered by Luan et al. The results show that, in terms of mean squared error for the chromatographic prediction of retention index, the non-linear model SVM is superior to the linear model HM (Liu et al., 2019). Shadrin et al. more recently presented their research comparing the uses of ANN and SVM for the prediction of environmental pollutants. The performance standards of the models were validated using the RMSE and the goodness of fit R2. The outcomes demonstrated that both non-linear models can reproduce the physical and biological effects of the samples (Shadrin, Pukalchik, Kovaleva, & Fedorov, 2020).

The use of artificial intelligence (AI)-based models in chemometrics design and chemical process modeling is superior and shows promise, according to prior research. Non-linear models, however, may have several problems, including overfitting, local minima, and generalizability, to mention a few. Numerous scientists claim that no one model is the best in terms of performance indices in the same or different data sets. Finding broadly applicable AI-based methods that may be used at numerous local scales is therefore essential. This study is the first in the technical literature to use the HPLC-DAD technique to measure and analyze the TQ content of black cumin from various geographical regions, including India, Syria, Turkey, Iraq, and Saudi Arabia. This study is also unique in that it is the first to compare the TQ content of black cumin from these regions to that of Saudi Arabia. Second, to the best of the authors' knowledge, this is the first work to show how these ensemble techniques may be used for TQ simulation using the HPLC-DAD technique.

In this study, the HPLC-DAD technique was used to simulate both the qualitative and quantitative properties of TQ using five different single models: principal component regression (PCR), least square-support vector machine (LSSVM), least square boost (LSQ-BOOST), adaptive neuro-fuzzy inference system (ANFIS), and step-wise linear regression (SWLR). Then, four innovative ensemble machine learning techniques—simple average ensemble (SAE), weighted average ensemble (WAE), neural network ensemble (NNE), and adaptive neuro-fuzzy inference system ensemble—were used to enhance the single models' predictive power (ANFIS-E). The mobile phase, which uses an isocratic elution system made up of de-ionized (D.I) water, methanol, and 2-propanol (Mp-A: Mp-B: Mp-C), as well as the concentration of the aqueous standard and flow rate, are thus the independent variables. Retention time (tR), one of the analyte's qualitative chromatographic qualities, and peak area (PA), one of the bioactive compound's quantitative chromatographic properties, are thought of as the study's outcome variables.

This study aims to investigate several models for TQ prediction and to compare and contrast the non-linear AI-based models with the conventional linear model. Additionally, to compare and contrast the ensemble models' performances and illustrate how they improve and boost the performance effectiveness of the individual models.

MATERIALS AND METHODS

Materials

The de-ionized water, methanol, isopropanol, ethanol, and TQ standard utilized in this work were all HPLC-grade chemicals that were purchased from Sigma Aldrich.

Instrumentation

An Agilent Technologies 1200 series HPLC instrument with a diode array detector (DAD) was used for the analysis. An Eclipse XDB-C18 (150 mm 4.6 mm, 5 m) reversed phase column was used to calculate the analyte TQ. For the mobile phase, an isocratic elution system made of de-ionized (D.I) water, methanol, and 2-propanol (Mp-A, Mp-B, and Mp-C) is used. The ideal flow rate was discovered to be 0.9 mL min-1 with an injection volume of 20 mL. 254 nm was chosen as the analytical wavelength and 16 minutes was chosen as the analysis time. The determination of the analyte was done by comparing the retention time of the pure standard with that of the real samples.

Sample preparation

The black cumin seeds were gathered from six distinct geographic areas: India, Syria, Turkey, Saudi Arabia, and Iraq. They were given as gifts by undergraduate students who returned to class in the winter following the summer break. All products are packaged and include the farm information on which they are produced. Particular attention was paid to ensuring that the samples were grown in the countries where they were taken. The seeds were further dried, grounded, powdered, and weighed. The material was then extracted with 100 mL of methanol and agitated for 2 hours using a magnetic stirrer. A rotary evaporator was used to evaporate the obtained extract. After that, the residue was diluted with ethanol and filtered through a solid phase extraction (SPE) cartridge (C8) in preparation for HPLC analysis (Isik, Kartal, & Erdem, 2017).

HPLC Quantification

TMQ solutions with ten distinct concentrations between 1 and 1000 ppm were generated for the quantitative analysis, and the peak areas of these concentrations were used for calibration.

Chemometrics and Models Conceptualization

Chemometrics are applied in two separate circumstances in this study. First, five alternative AI-based models, including SWLR, PCR, LSSVM, ANFIS, and LSQ-Boost employing various independent variables, were used to simulate the chromatographic behavior of TQ informing of retention duration and peak area. Second, SAE, WAE, NNE, and ANFIS-E are informed by the application of ensemble machine learning to enhance the performance of AI-based models.

Using a variety of concentrations of the analyte's standard solution, flow rate, and a mobile phase made up of de-ionized (D.I.) water, methanol, and 2-propanol, the chromatographic behavior of TQ in terms of retention time and peak area was modeled.

Phase 1: Data Acquirement

The experimental studies, which were based on the calibration of a standard TQ solution, produced the entire data set. Additionally, the data points were split into two groups: 30% for the testing stage and 70% for the calibration stage. Following data validation, potential modeling issues like overfitting and underfitting were checked and controlled (Abba et al., 2020).

Phase 2: Data normalization, statistical analysis, and correlation

Based on equation 1, the input and output factors utilized in this investigation were both standardized into a range of 0 to 1. Before modeling, normalization lowers data redundancy and decreases significant numerical errors, which is one of its main benefits.

$$y = \left(\left(\frac{x - x_{min}}{x_{max} - x_{min}}\right)\right) \tag{1}$$

Excel 2016 was used for the statistical analysis and correlation, and a 95% confidence level was used.

Phase 3: Simulation using single models

MATLAB 9.3 was used to run the individual AI-based models (PCR, LSSVM, ANFIS, and LSQ-Boost) as well as the conventional linear regression SWLR (R2020a).

Phase 4: Ensemble machine learning techniques

To increase the performance effectiveness of the individual models, ensemble machine learning approaches (PCR, LSSVM, ANFIS, LSQ-Boost, and SWLR) assemble, add, and integrate both linear and non-linear models.

Phase 5: Performance evaluation metrics

The performance evaluation parameters for any type of data-driven technique are examined using a variety of criteria by contrasting experimental and simulated values. Four separate performance evaluation criteria were used to compare the performance of the individual models and the ensemble machine-learning approaches established in this work during the calibration and verification phases.

Equations 2 through 5 are used to calculate the mean square error (MSE), root mean square error (RMSE), correlation coefficient (CC), and Nash-Sutcliffe efficiency (NSE), respectively.

$$MSE = \frac{1}{N}\Sigma_{i=1}^{N}\left(Y_{obsi} - Y_{comi}\right)^2 \tag{2}$$

$$RMSE = \sqrt{\frac{\Sigma_{i=1}^{N}\left(Y_{obsi} - Y_{comi}\right)^2}{N}} \tag{3}$$

$$CC = \frac{\Sigma_{i=1}^{N}\left(Y_{obs} - \overline{Y}_{obs}\right)\left(Y_{com} - \overline{Y}_{com}\right)}{\sqrt{\Sigma_{i=1}^{N}\left(Y_{obs} - \overline{Y}_{obs}\right)^{2}\Sigma_{i=1}^{N}\left(Y_{com} - Y_{com}\right)^{2}}} \quad (4)$$

$$NSE = 1 - \frac{\Sigma_{j=1}^{N}\left[(Y)_{obs,j} - (Y)_{com,j}\right]^{2}}{\Sigma_{j=1}^{N}\left[(Y)_{obs,j} - \overline{(Y)}_{obs,j}\right]^{2}} \quad (5)$$

Table 1. The amount of thymoquinone in Nigella sativa seed extracts from different regions

Sample Name	Concentration (ppm)	% Thymoquinone amount
Iraq	31.26 ± 0.071	0.031
India	30.35 ± 0.167	0.030
Saudi Arabia	22.15 ± 0.165	0.022
Syria	5.47 ± 0.099	0.005
Turkey	1.39 ± 0.112	0.001

RESULTS

HPLC Linearity Results

By creating a calibration curve, the standard solutions of TQ were prepared between 0 and 1000 ppm concentration.

Our investigation revealed that the samples of black cumin seeds from various geographic locations with the highest concentration of TQ were those from Iraq and India (Figure 1). TQ was present in all samples, but in varying amounts; the amounts of TQ in the samples from Iraq, India, Saudi Arabia, Syria, and Turkey, respectively, were 0.031, 0.030, 0.022, 0.005, and 0.001%. The soil quality, climate changes in the places where the seeds are cultivated, and other factors are thought to be responsible for the variations in the TQ of black cumin seeds. The amount of TQ discovered in *Nigella sativa* seeds obtained from Ankara was discovered to be in the range of 0.010-0.376% due to changes in parameters like those described in a previous study by our team (Isik et al., 2017). The amount of TQ discovered in the seeds of the black cumin plant grown in Kuwait and India ranged from 1039.85 mg/kg to 2940.43 mg/kg (Herlina, Aziz, Kurniawati, & Faridah, 2017). Black cumin seeds were discovered to contain 0.06% TQ by Gholamnezhad et al. (2015) in a study looking into the immunomodulatory and cytotoxic effects of the seeds. Although the TQ range identified in the literature is consistent with the amount of TQ computed within the scope of this study, it can be argued that the amount of TQ is in a very wide range. This vast range of TQ in plants is influenced by genetic abnormalities, harvest period/season, and physiological circumstances (Zribi, Omezzine, & Haouala, 2014). To provide a standard effect and concentration of its main analyte, TQ, it is crucial to ascertain the phytochemical composition of the *Nigella sativa* plant that will be employed for therapeutic purposes.

Using the established calibration equation, the levels of thymoquinone identified in various black cumin seeds from various geographical regions are estimated and summarized in Table 1.

Performance of the chemometrics-based models

The quantitative performance effectiveness of the individual models PCR, LSSVM, ANFIS, LSQ-Boost, and SWLR is displayed in Table 2. The ANFIS model outperforms all the other single models (PCR, LSSVM, LSQ-Boost, and SWLR) in modeling both tR and PA in the training and testing stages, according to the comparative performance of these techniques (see Table 2). According to Nourani et al. 2012, any model must have a minimum Nash-Sutcliffe efficiency (NSE) of 80% to be accepted. (In other words, the model needs to have a minimum R2-value of 0.8 to be considered acceptable) (Nourani, Hakimzadeh, & Amini, 2012). Only ANFIS, according to the performance table, was able to imitate tR, which is mostly attributable to its intricate capacity to model highly non-linear data. This is consistent with the findings of our earlier investigations (Usman, Işik, & Abba, 2021; Abdullahi Garba Usman, Işik, Abba, & Meriçli, 2021a ; and Abdullahi Garba Usman, Isik, Abba, & Mericli, 2021b). In contrast, only SWLR, LSSVM, and ANFIS were able to meet the minimal 80% threshold for a model to be acceptable when it came to the simulation of PA. This demonstrated the necessity for additional methods, such as ensemble machine learning, to improve the performance of the individual models.

The effectiveness of the single models (PCR, LSSVM, ANFIS, LSQ-Boost, and SWLR)

The scatter plot of the comparative performance of the models is illustrated in Figure 2.

According to a comparison of their performances, the four ensemble machine learning techniques can simulate the chromatographic properties of TQ, PA, and tR with minimum and maximum NSE-values of 0.842 and 0.999 in the training phase and 0.918 and 0.999 in the testing phase, respectively. Comparing the approaches to the individual models, they both demonstrated lower error values in terms of RMSE and MSE.

Performance of the Ensemble machine learning techniques

Table 3 provides a summary of the data from ensemble machine learning. To better understand the ensemble machine learning algorithms' exploratory performance for the prediction of tR and PA, scatter plots are used (see Figure 3).

A new graphic design tool dubbed the "Taylor diagram" was used to demonstrate the performance capabilities of the ensemble machine learning technique. Due to its importance, this diagram has been applied to numerous modeling fields, including computer science, computer vision, and climate modeling.

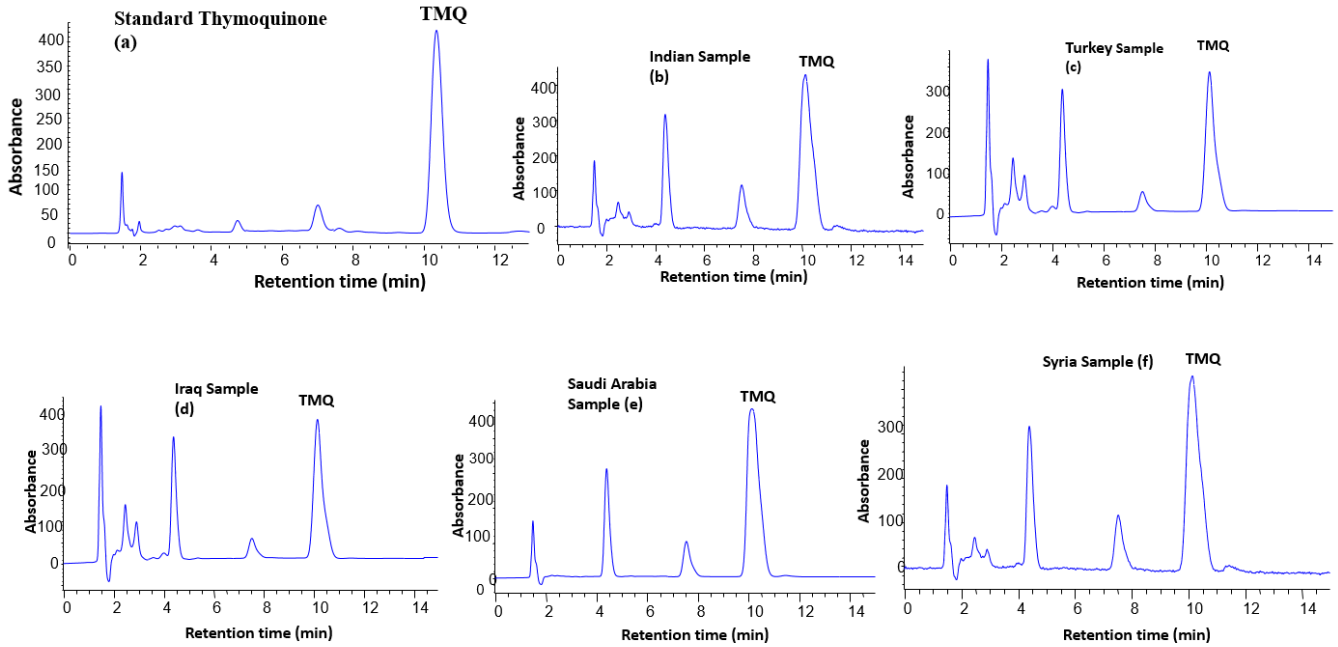

Figure 1. The chromatogram of the a) standard TQ analyte, b) Indian c) Turkey (d) Iraq sample e)Saudi Arabia, and f) Syrian sample

Table 2. Performance table of the single models (PCR, LSSVM, ANFIS, LSQ-Boost, and SWLR)

	Training				Testing			
	NSE	**CC**	**RMSE**	**MSE**	**NSE**	**CC**	**RMSE**	**MSE**
SWLR-tR	0.362	0.601	0.192	0.037	0.220	0.469	0.004	0.000
PCR-tR	0.701	0.837	0.131	0.017	0.586	0.766	0.003	0.000
LSSVM-tR	0.173	0.416	0.218	0.048	0.016	0.128	0.005	0.000
ANFIS-tR	0.999	1.000	0.001	0.000	0.999	1.000	0.000	0.000
LSQ-Boost-tR	0.613	0.783	0.149	0.022	0.651	0.807	0.003	0.000
SWLR-PA	0.999	1.000	0.005	0.000	0.987	0.994	0.009	0.000
PCR-PA	0.674	0.821	0.126	0.016	0.234	0.483	0.070	0.005
LSSVM-PA	0.923	0.961	0.022	0.000	0.983	0.991	0.029	0.001
ANFIS-PA	0.999	1.000	0.000	0.000	0.999	1.000	0.005	0.000
LSQ-Boost-PA	0.765	0.875	0.039	0.002	0.567	0.753	0.146	0.021

The graphic is used to demonstrate the study's findings regarding the goodness of fit in terms of CC (see Figure 4).

Table 3. Performance of the Ensemble machine learning techniques (SAE, WAE, NNE, and ANFIS-E)

	Training				Testing			
	NSE	CC	RMSE	MSE	NSE	CC	RMSE	MSE
SAE-tr	0.842	0.917	0.087	0.009	0.843	0.918	0.002	0.000
WAE-tr	0.844	0.919	0.086	0.009	0.890	0.944	0.002	0.000
NNE-tr	0.989	0.995	0.025	0.001	0.986	0.993	0.001	0.000
ANFIS-E-tr	0.999	1.000	0.000	0.000	0.999	0.999	0.000	0.000
SAE-PA	0.998	0.999	0.010	0.000	0.987	0.993	0.025	0.001
WAE-PA	0.999	0.999	0.008	0.000	0.998	0.999	0.009	0.000
NNE-PA	0.999	1.000	0.015	0.000	0.999	1.000	0.005	0.000
ANFIS-E-PA	0.999	1.000	0.001	0.000	0.999	0.999	0.004	0.000

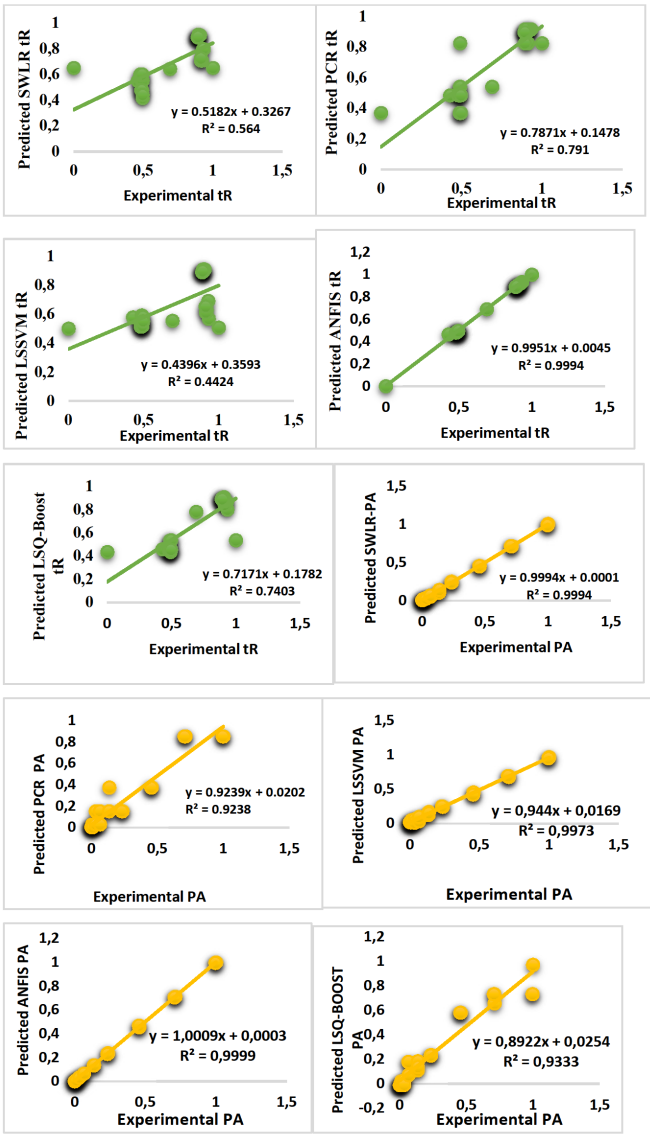

Figure 2. Scatter plots of the single models for their respective retention time (tR) and Peak area (PA).

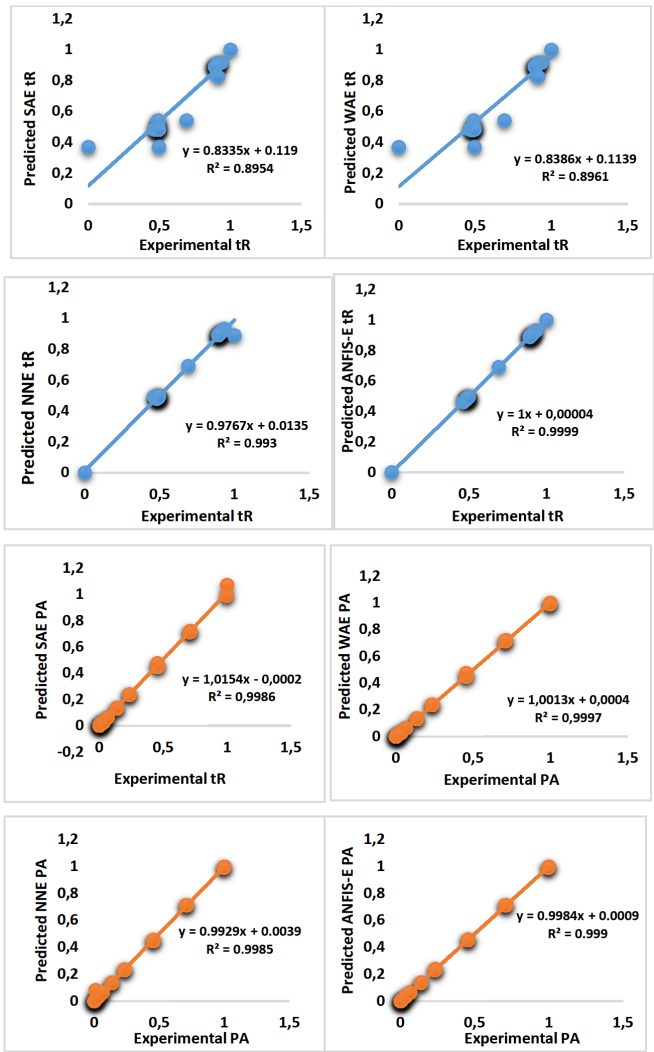

Figure 3. Scatter plots of the ensemble techniques tR and for PA

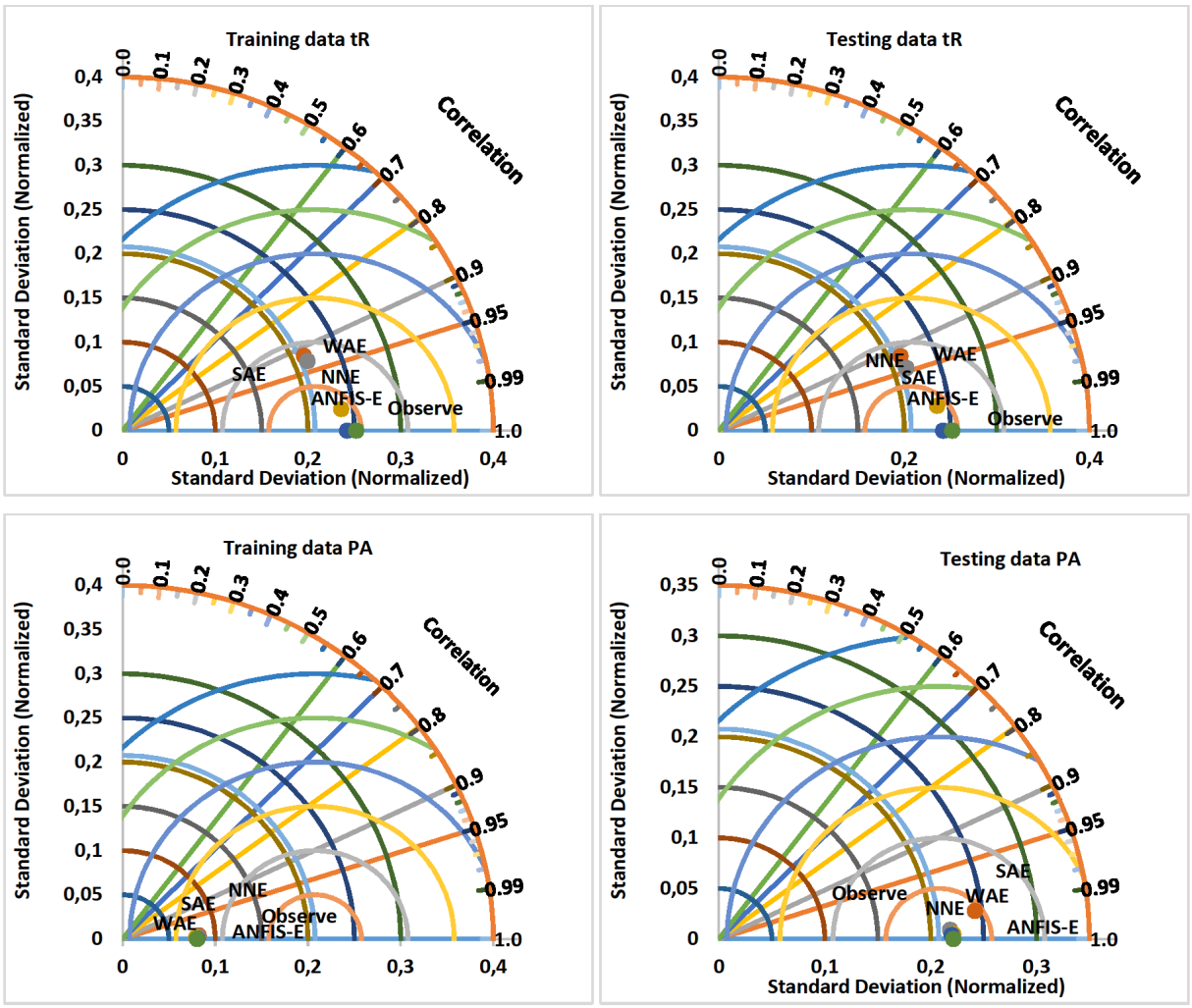

Figure 4. Taylor diagrams of the ensemble techniques tR and for PA

CONCLUSION

To measure the TQ content of each sample of black cumin collected from various geographical locations, a quantitative analysis of each sample was carried out. The overall amount of TQ contained in each sample is influenced by a variety of geographical conditions, including rainfall, seasonal changes, and soil, as mentioned in the literature. The collected results demonstrated that samples from India and Iraq have the highest concentration of TQ. TQ was present in all samples, but in varying amounts; the amounts of TQ in the samples from Iraq, India, Saudi Arabia, Syria, and Turkey, respectively, were 0.031, 0.030, 0.022, 0.005, and 0.001%.

As one of the most modern chemometrics techniques for simulating the chromatographic behavior of various analytes, the study also included the deployment of both single models and ensemble machine-learning methodologies. The chemometrics results show that these models are capable of simulating both the qualitative and quantitative features of TQ.

In addition, additional metaheuristic methods can be employed to simulate the chromatographic characteristics of TQ, such as particle swam optimizations (PSO) and Harris Hawks optimization techniques (HHO).

Peer Review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- S.I., A.G.U., S.I.A.; Data Acquisition- S.I., A.G.U.; Data Analysis/Interpretation- S.I., A.G.U., S.I.A.; Drafting Manuscript- S.I., A.G.U.; Critical Revision of Manuscript- S.I., A.G.U., S.I.A.; Final Approval and Accountability- S.I., A.G.U., S.I.A.

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: The authors declared no financial support.

ORCID IDs of the authors

Selin Isık	0000-0001-7601-3746
Abdullahi Garba Usman	0000-0001-5660-4581
Sani Isah Abba	0000-0001-9356-2798

REFERENCES

Abba, S. I., Pham, Q. B., Usman, A. G., Linh, N. T. T., Aliyu, D. S., Nguyen, Q., & Bach, Q. V. (2020). Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *Journal of Water Process Engineering*, 33, 101081. doi: 10.1016/j.jwpe.2019.101081

Chen, M., Wen, F., Zhang, Y., Li, P., Zheng, N., & Wang, J. (2019). Determination of native lactoferrin in milk by HPLC on HiTrapTM Heparin HP column. *Food Analytical Methods*, 12(11), 2518-2526. doi: 10.1007/s12161-019-01572-x

Golmohammadi, H., Dashtbozorgi, Z., & Vander Heyden, Y. (2015). Support Vector Regression Based QSPR for the Prediction of Retention Time of Peptides in Reversed-Phase Liquid Chromatography. *Chromatographia*, 78(1-2), 7-19. doi: 10.1007/s10337-014-2819-1

Herlina, Aziz, S. A., Kurniawati, A., & Faridah, D. N. (2017). Changes of Thymoquinone, Thymol, and Malondialdehyde Content of Black Cumin (*Nigella sativa L.*) in Response to Indonesia Tropical Altitude Variation. *HAYATI Journal of Biosciences,* 24(3), 156-161. doi: 10.1016/j.hjb.2017.08.004

Isik, S., Kartal, M., & Erdem, S. A. (2017). Quantitative analysis of thymoquinone in *Nigella sativa L.* (Black Cumin) seeds and commercial seed oils and seed oil capsules from Turkey. *Ankara Universitesi Eczacilik Fakultesi Dergisi*, 41(1), 34-41. doi: 10.1501/Eczfak_0000000593

Liu, Y., Kohlberger, T., Norouzi, M., Dahl, G. E., Smith, J. L., Mohtashamian, A., Stumpe, M. C. (2019). Artificial intelligence–based breast cancer nodal metastasis detection insights into the black box for pathologists. *Archives of Pathology and Laboratory Medicine*, 143(7), 859-868. doi: 10.5858/arpa.2018-0147-OA

Marrero-Ponce, Y., Barigye, S. J., Jorge-Rodríguez, M. E., & Tran-Thi-Thu, T. (2018). QSRR prediction of gas chromatography retention indices of essential oil components. *Chemical Papers*, 72(1), 57-69. doi: 10.1007/s11696-017-0257-x

Nourani, V., Hakimzadeh, H., & Amini, A. B. (2012). Implementation of artificial neural network technique in the simulation of dam breach hydrograph. *Journal of Hydroinformatics*, 14(2), 478. doi: 10.2166/hydro.2011.114

Rezai, F., Işık, S., Kartal, M., & Aslan Erdem, S. (2018). Effect of priming on thymoquinone content and in vitro plant regeneration with tissue culture of black cumin (*Nigella sativa L.*) seeds. *Journal of Chemical Metrology*, 12(2), 89-98. doi: 10.25135/jcm.18.18.09.950

Shadrin, D., Pukalchik, M., Kovaleva, E., & Fedorov, M. (2020). Artificial intelligence models to predict acute phytotoxicity in petroleum contaminated soils. *Ecotoxicology and Environmental Safety*, 194(February), 110410. doi: 10.1016/j.ecoenv.2020.110410

Tham, S. Y., & Agatonovic-Kustrin, S. (2002). Application of the artificial neural network in quantitative structure-gradient elution retention relationship of phenylthiocarbamyl amino acids derivatives. *Journal of Pharmaceutical and Biomedical Analysis*, 28(3-4), 581-590. doi: 10.1016/S0731-7085(01)00690-2

Usman, A. G., Işik, S., & Abba, S. I. (2021). Hybrid data-intelligence algorithms for the simulation of thymoquinone in HPLC method development. *Journal of the Iranian Chemical Society,* 18(7), 1537-1549. doi: 10.1007/s13738-020-02124-5

Usman, Abdullahi Garba, Isik, S., Abba, S. I., & Mericli, F. (2021a). Artificial intelligence based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method. *Turkish Journal of Chemistry*, 44(5). doi: 10.3906/kim-2003-6

Usman, Abdullahi Garba, Işik, S., Abba, S. I., & Meriçli, F. (2021b). Chemometrics-based models hyphenated with ensemble machine learning for retention time simulation of isoquercitrin in Coriander sativum L. using high-performance liquid chromatography. *Journal of Separation Science*, 44(4), 843-849. doi: 10.1002/jssc.202000890

Vasiljević, T., Onjia, A., Čokeša, D., & Laušević, M. (2004). Opti-

mization of artificial neural network for retention modeling in high-performance liquid chromatography. *Talanta*, 64(3), 785-790. doi: 10.1016/j.talanta.2004.03.032

Zribi, I., Omezzine, F., & Haouala, R. (2014). Variation in phytochemical constituents and allelopathic potential of *Nigella sativa* with developmental stages. South *African Journal of Botany*, 94, 255-262. doi: 10.1016/j.sajb.2014.07.009

How cite this article

Isik, S., Usman, A.G., Abba, S.I. (2023). A chemometrics-based approach for the determination of thymoquinone from Nigella sativa L. (Black Cumin) seeds of different geographical regions using the HPLC technique. *İstanbul Journal of Pharmacy, 53*(3), 320-328. DOI: 10.26650/IstanbulJPharm.2023.1181298