

Full Length Article

A comparative study of machine learning algorithms trained with Monte Carlo simulations for X-ray fluorescence analysis

Kaan Yavaş^a, Ulaş Keskin^b, Ozan Toker^c, Özgür Akçalı^c, H. Birtan Kavanoz^c, Orhan İçelli^{c,*}^a Istanbul Gelisim University, Faculty of Engineering and Architecture, Department of Basic Sciences, Istanbul, Turkey^b Bursa Uludağ University, Faculty of Arts and Sciences, Department of Physics, Bursa, Turkey^c Yıldız Technical University, Faculty of Arts and Sciences, Department of Physics, Istanbul, Turkey

ARTICLE INFO

Keywords:

X-ray fluorescence
Machine learning
Artificial neural network
Random forest
Support vector regression
Quantitative analysis

ABSTRACT

X-ray fluorescence (XRF) spectrometry is widely used in quantitative analysis because it is fast, reliable, and non-destructive. However, challenges such as matrix effects, spectral overlap, background interference, and the need for suitable standards can make quantitative XRF analysis difficult, motivating the use of machine learning models that learn composition spectrum relationships from representative datasets. Normalized characteristic peak areas extracted from XRF spectra were used as input to predict the elemental compositions of stainless steel samples. Artificial neural network (ANN), random forest (RF), and support vector regression (SVR) algorithms were trained with datasets generated from Monte Carlo simulations. All trained models were validated using experimental XRF spectra. All models demonstrated compatibility with experimental results, with the ANN emerging as the most effective model, particularly for trace elements. These results demonstrate the feasibility of using simulation-generated datasets to train machine learning models and reduce reliance on extensive experimental data.

1. Introduction

Stainless steels are industrial materials used in many areas such as automation and construction [1–5]. The elemental composition of stainless steel directly influences its strength, toughness, and resistance to chemical and environmental factors. For example, chromium enhances corrosion resistance, nickel improves ductility, molybdenum increases resistance to pitting corrosion, and iron acts as the base element [6,7]. Therefore, determining both the stainless-steel type and its elemental distribution is crucial for assessing its suitability for a given application.

X-ray fluorescence (XRF) spectrometry is widely preferred for analyzing stainless steel and other materials in fields such as archaeology, mining, and biology due to its speed, reliability, and non-destructive nature [8–10]. Despite its advantages, XRF faces challenges such as matrix effects, the need for standard samples, overlapping peaks, and background interference. These challenges complicate quantitative analysis and often require the application of analytical, empirical, or semi-empirical methods, either individually or in

combination [11–13]. Therefore, in addition to analytical and empirical methods, Monte Carlo-based simulation codes are widely used to improve the accuracy and precision of XRF analysis.

Monte Carlo-based simulation codes such as Monte Carlo N-Particle (MCNP), GEANT, and FLUKA are widely used in fields such as radiation-matter interactions, including reactor physics, medical physics, and radiation protection [14–17]. These codes are also used to address challenges in XRF spectrometry and to improve quantitative analysis [18–21]. Furthermore, researchers have developed various methods for performing quantitative analysis using Monte Carlo simulations [22–25].

Machine learning algorithms are currently applied across a wide range of fields [26,27]. Previous studies have reported promising results for the use of machine learning in XRF analysis, although the reported performance varies depending on the application. Bos et al. [28] applied ANN-based multivariate calibration to the Fe–Cr–Ni system in XRF spectroscopy and showed that neural networks could model nonlinear interelement effects and provide competitive performance over broad concentration ranges. Allegretta et al. [29] combined pED-XRF spectra

* Corresponding author.

E-mail address: oiçelli@yildiz.edu.tr (O. İçelli).<https://doi.org/10.1016/j.nimb.2026.166289>

Received 7 January 2026; Received in revised form 22 May 2026; Accepted 14 August 2026

Available online 24 August 2026

0168-583X/© 2026 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.



Fig. 1. Experimental setup.

with PCA and supervised machine learning algorithms for the macro-classification of meteorites and meteor-wrongs, obtaining 100% classification accuracy for selected classifiers when standardized main-range spectra were used. Jones et al. [30] trained a CNN using synthetic pigment XRF spectra generated with the fundamental parameters method. The model reached 100% accuracy on synthetic spectra, but its accuracy decreased to 55% on real reference spectra and increased to 96% after transfer learning with a limited set of real spectra. This result highlights both the usefulness of synthetic spectra for model training and the importance of minimizing the mismatch between simulated and experimental spectra. For quantitative XRF analysis, Rakotondrajoa and Radtke [31] trained ANN models with Monte Carlo-generated XRF spectra and reported mean absolute errors of 0.1 wt% for synthetic data and 0.7 wt% for experimental certified reference materials [31]. Lee et al. [32] used raw XRF scanning spectra to quantify CaCO_3 and TOC in marine sediments and reported test-set R^2 values of 0.96 and 0.78, respectively. Zhao et al. [33] developed a physics-based ED-XRF spectra simulator for data enhancement and used simulated spectra as input to an XGBoost model, reporting high R^2 values for simulated elemental concentration ranges [33]. Taken together, these studies suggest that machine learning can support both classification and quantification in XRF analysis, provided that the training data sufficiently represents the spectral and matrix-related features of the target measurements. However, because previous studies differ in sample matrix, spectral simulation strategy, preprocessing, model type, and validation design, a systematic comparison of different machine learning algorithms trained on MCNP-simulated stainless-steel XRF spectra is still needed.

The aim of this study is to determine the elemental composition of stainless steel using XRF spectra through machine learning algorithms trained exclusively on MCNP simulations. A simulation geometry consistent with the experimental XRF setup was designed using the MCNP code. Training and validation datasets were created by normalizing the characteristic peaks of spectra obtained from simulations and experimental measurements. The machine learning algorithms were trained using datasets generated from MCNP simulations, whereas experimental datasets were used only for model validation.

2. Materials and methods

2.1. Experimental setup

AMPTEK Mini-X X-ray tube source rated at 40 kV/100 μA was operated at 35 kV to obtain characteristic X-rays of elements. AMPTEK X-123 silicon drift detector (SDD) with 122 eV and 129 eV resolution in the range of 1–40 keV was used to count characteristic X-rays. X-ray tube

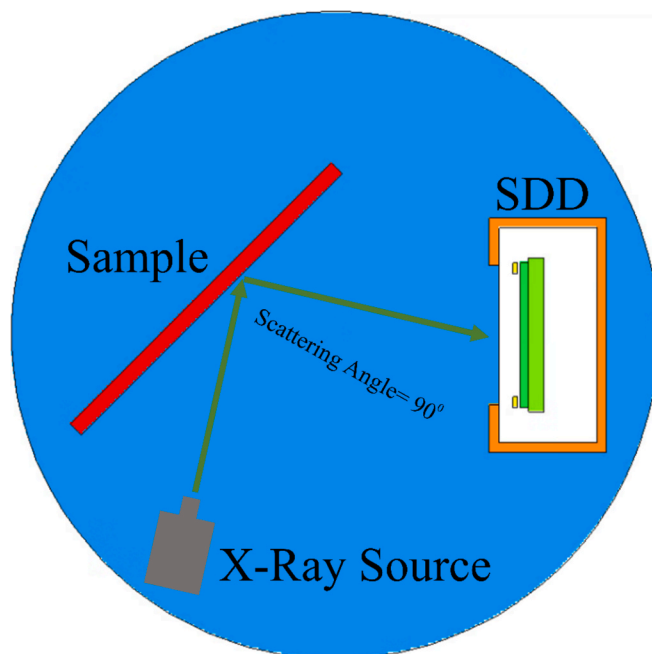


Fig. 2. Simulation geometry used in MCNP.

Table 1

Range and increment of percentage weights of elements in stainless steel used in simulations.

Element	Cr	Ni	Fe	Mo
Weight Range (%)	12–26	0–18	Fill	0–4
Increment	2	2	–	1

and SDD were used in conjunction with an automated system designed to minimize experimental errors. This system precisely adjusts the scattering angle and the distance between the sample, the tube, and the SDD, ensuring optimal alignment and consistent setup for accurate measurements. Fig. 1 shows the experimental setup. The sample spectra were used only after the machine learning algorithms had been trained. This allowed the compatibility of the models with the experimental data to be evaluated.

2.2. MCNP geometry

In this study, the experimental setup was modeled in MCNP to obtain stainless-steel spectra. Fig. 2 shows the MCNP geometry. The energy distribution of the X-ray tube operated at 35 kV was used as a point source. To obtain detailed and experimentally comparable XRF spectra, spectra were scored using an F8 pulse-height tally with 0.05 keV energy bins over the 1–20 keV range.

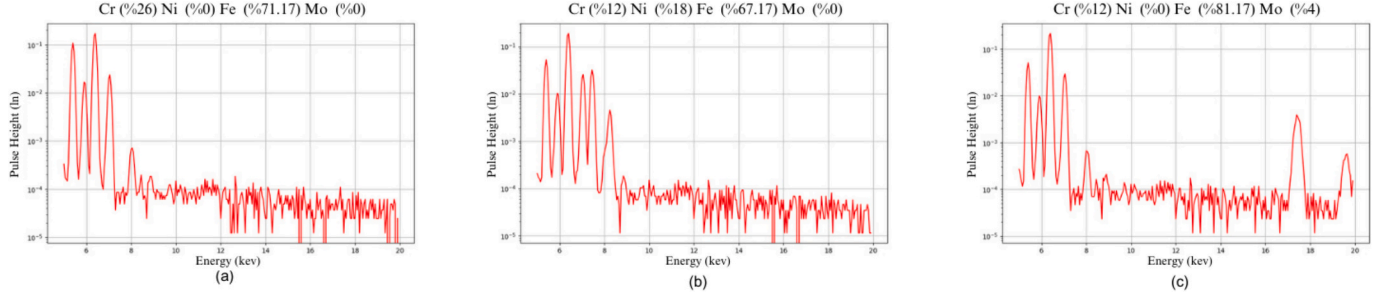
To generate a dataset for machine learning algorithms, a total of 400 different alloy combinations were created using Cr, Fe, Ni, and Mo with various intervals and increments together with fixed matrix elements. Alloy combinations are given in Table 1, the fixed matrix elements are given Table 2. XRF spectra for each combination were obtained using MCNP simulations.

Previous studies have evaluated the performance of the EPRDATA12 library in X-ray simulation applications. Cook [34] compared EPRDATA12 with the earlier MCLIB84 library and reported that EPRDATA12 provided a more accurate description of characteristic X-ray production [34]. Poskus [35] showed that the single-event model implemented in EPRDATA12 gives more reliable results for characteristic X-ray generation than previous condensed-history approaches. Jung et al. [36] demonstrated that MCNP6.1 calculations using

Table 2

Fixed matrix elements of stainless steels used in simulations.

Element	C	N	Si	P	S	V	Mn	Co	Cu	Nb	Ag	W
Weight (%)	0.061	0.015	0.55	0.029	0.026	0.062	1.46	0.126	0.437	0.011	0.1	0.045

**Fig. 3.** Representative spectra generated from MCNP simulations.

EPRDATA12 were in close agreement with GATE and PENELOPE, indicating that this library is suitable for Monte Carlo simulations involving low-energy photon-induced fluorescence X-rays and Auger electrons [36]. Parsons [37] reported that the differences between EPRDATA12 and the more recent EPRDATA14 library were not significant [37]. Jung [38] also applied MCNP6.1 with EPRDATA12 to determine detection limits in both L-shell and K-shell XRF imaging systems, confirming its practical applicability to XRF-related simulations. Overall, these studies indicate that EPRDATA12 can be considered a reliable data library for X-ray simulation studies [38]. For this reason, EPRDATA12 was used in the present study to generate the characteristic X-rays of the materials [39].

2.3. Data analysis

XRF spectra were obtained from the simulation of 400 different steels created with the values given in Table 1. Examples of spectra obtained with MCNP simulations are shown in Fig. 3. The peak area (N_i) of the characteristic X-ray for each element was normalized to obtain the normalized peak areas (n_i) according to the formula given below

$$n_i = \frac{N_i}{\sum_{i=1}^m N_i} \quad (1)$$

where m is the number of analyte elements, N_i is the raw peak area of the i^{th} element, and n_i is its corresponding normalized value. The dataset was divided into 70% training and 30% test sets for each algorithm. Experimental spectra were processed using the same normalization procedure to create the validation dataset.

2.4. Machine learning

Researchers in many fields such as mathematics, physics and engineering use regression analysis to reveal the correlation between variables or to predict an unobserved value in statistical modeling [26,40]. In machine learning algorithms, various mathematical methods are employed in regression analysis, such as linear regression, logistic regression, and polynomial regression. While some of these methods can only address linear relationships between variables, others do not allow for multi-regression analysis that simultaneously examines multiple independent variables [40].

In this study, we used Support Vector Regression (SVR), Random Forest (RF) and ANN to perform quantitative analysis of stainless steel. These models were implemented in a Python (v3.12) environment [41].

Table 3

Hyperparameter optimization range for the SVR model.

Hyperparameter	Range of Evaluated Values
Kernel $\theta(x)$	'linear', 'rbf', 'poly', 'sigmoid'
C (Regularization Parameter)	0.01, 0.1, 1, 10, 100, 1000, 10,000
Loss Function (ϵ)	0.001, 0.01, 0.1, 0.5, 1, 2
Gamma	'scale', 'auto', 0.001, 0.01, 0.1, 1, 10
Degree	2, 3, 4, 5
Shrinking	True, False
Cache Size	200, 500, 1000

Table 4

Hyperparameter optimization range for the RF model.

Hyperparameter	Range of Evaluated Values
Maximum Depth (max_depth)	5, 10, 15, 30, 50, None
Min Samples Split	2, 3, 5, 7, 10
Min Samples Leaf	1, 2, 3, 4, 5
Number of Trees (n_estimators)	100, 200, 300, 500, 1000, 3000, 5000
Maximum Features (max_features)	'sqrt', 'log2', 2, 3, 4
ccp_alpha	0.0, 0.0001, 0.001

2.4.1. Support vector regression

SVR is a support-vector-machine-based model used to solve linear and nonlinear regression problems [42]. SVR operates by constructing a hyperplane in a multidimensional feature space while allowing a specified error margin. It aims to minimize prediction error within the defined margin while controlling model complexity [43]. In this work, the SVR model was implemented using the Scikit-learn (v1.3.0) library [44]. To determine the most robust model for predicting elemental concentrations, a comprehensive grid search was performed across 28,224 different hyperparameter combinations. The extensive search space investigated during this process is detailed in Table 3.

2.4.2. Random Forest

RF is used for classification and regression by combining multiple decision trees into an ensemble model [45]. Various decision tree algorithms, such as ID3, C4.5, CART, and CHAID, can be used to construct these trees [46]. RF predictions are obtained by averaging the outputs of multiple regression trees (T), which are built using input vectors comprising $\times$ different features extracted from the observation dataset [47]. The RF model was applied via the Scikit-learn framework [44]. The optimization process for the RF model involved a systematic evaluation of 15,750 unique combinations of hyperparameters, including tree depth and the number of estimators, to achieve optimal predictive

Table 5

Hyperparameter optimization range for the ANN model.

Hyperparameter	Range of Evaluated Values
Hidden Units	16, 32, 64, 128, 256, 512, 1024
Dropout Rate	0.1, 0.2, 0.3
Learning Rate	0.001, 0.0005
Weight Decay	0.00001, 0.0001, 0.001, 0.01
Epochs	50, 100, 250, 500, 1000
Optimizer	adam, SGD, rmsprop

performance. These parameters are summarized in Table 4.

2.4.3. Artificial neural networks (ANN)

Artificial Neural Networks (ANN) are among the most commonly used algorithms for nonlinear classification and regression tasks. In ANN, neurons are organized to enable a one-way flow of information from the input layer to the output layer through hidden layers and the model is trained using backpropagation [48]. The neurons in each layer are interconnected via weight values, which are optimized by the ANN algorithm. The number of neurons and layers can be adjusted based on the complexity of the problem being analyzed [43,49]. In this research, the ANN architecture was constructed and trained using the PyTorch (v2.1) library [50]. The architecture of the ANN was optimized by exploring 2520 distinct hyperparameter sets, encompassing different hidden-unit numbers, dropout rates, learning rates, weight-decay values, epoch numbers, and optimizers. The detailed configuration of the architectural search space is provided in Table 5.

3. Result and discussion

Mean Squared Error (MSE) and R^2 metrics were employed to evaluate and compare the regression models. MSE was used to quantify the average squared difference between predicted and reference values, whereas R^2 was used as a coefficient-of-determination score to evaluate the model fit relative to the variance of the reference data [44,51]. The formulas for these performance metrics are provided below, where y represents the value predicted by the model.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (3)$$

SVR allows analysis of the relationship between only one dependent variable and independent variables at a time. Therefore, to examine each element, four separate SVR models were developed, in which the normalized peaks of the characteristic X-rays of all elements were chosen as independent variables, and the percentage weight of the target element was selected as the dependent variable. Using the generated dataset, hyperplanes were constructed in a four-dimensional feature space using the SVR model. Hyperparameter optimization was performed for each SVR model. During the hyperparameter evaluation process, MSE was used as the performance metric. The selected hyperparameters for the SVR models are presented in Table 6.

Since the RF algorithm allows multi-output regression, a single

Table 7

Optimal hyperparameters for random forest.

Hyperparameter	Optimal Value
Maximum Depth (max_depth)	15
Min Samples Split	2
Min Samples Leaf	1
Number of Trees (n_estimators)	1000
Maximum Features (max_features)	3
ccp_alpha	0.0

model was trained to predict all elements simultaneously. During the hyperparameter evaluation process, MSE was used as the performance metric. The selected hyperparameters for the RF model are presented in Table 7. Although increasing the number of trees reduced MSE, the error stabilized beyond a certain point. Fig. 4 shows the relationship between the number of trees and MSE. Similarly, the tree depth was observed to stabilize beyond a depth of 15. Fig. 5 illustrates the MSE as a function of tree depth.

The ANN algorithm was designed to predict all elements simultaneously. During the hyperparameter evaluation process, MSE values were used as the performance metric. The optimal hyperparameters for the ANN model are presented in Table 8. Fig. 6 shows the MSE for the training and test datasets of each element as a function of the number of neurons. After determining the optimal number of neurons, the MSE was evaluated based on the number of epochs. Fig. 7 illustrates the MSE as a function of the number of epochs.

The performance metrics of the trained models on the training dataset are presented in Table 9. For the training dataset, the MSE scores for all elements are as low as 10^{-6} for RF, approximately 10^{-5} for SVR, and in the range of 10^{-5} for ANN. Considering only the training dataset performance metrics, the SVR models have the lowest R^2 score, around 0.82. In contrast, the R^2 values for the ANN model are approximately 0.99 for all elements, while the RF model achieves R^2 values above 0.99 for Cr, Fe, and Ni, but a notably lower value of 0.95 for Mo, indicating that the RF model already shows limited learning capacity for Mo during training. All models exhibit MSE scores close to zero on the training dataset, indicating acceptable performance levels. Additionally, since the SVR algorithm requires training separate models for each element to be predicted, it can be inferred that the other two models have an advantage over SVR at this stage.

Metrics for the test dataset, evaluated using the simulation dataset, are presented in Table 10. For SVR, the MSE scores for all elements are in the range of 10^{-5} , while the RF algorithm scores are also on the order of 10^{-5} for all elements. For ANN, the MSE scores are approximately 10^{-4} for Cr and in the range of 10^{-5} for Fe, Ni, and Mo. The SVR model maintains an R^2 score of around 0.82 on the test dataset, consistent with its training performance. Although the RF model achieves an R^2 of ~ 0.99 for all elements on the training set, the test performance drops for Mo ($R^2 \approx 0.65$). This suggests overfitting for Mo despite hyperparameter optimization. This behavior should not be interpreted as the result of an insufficient hyperparameter search, since the RF model was optimized over a broad hyperparameter space including 15,750 unique hyperparameter combinations, and the values reported in Table 7 correspond to the optimum combination obtained from this search. Instead, the decrease in RF performance is mainly target specific and is most evident for Mo. Although an out-of-bag estimate was also considered as an

Table 6

Optimal hyperparameters for support vector regression.

Hyperparameter Model	C	Cache Size	Degree	Epsilon	Gamma	kernel	Shrinking
Model For Cr	1	200	3	0.01	scale	poly	True
Model For Fe	1	200	2	0.01	scale	Poly	True
Model For Ni	100	200	2	0.01	scale	linear	True
Model For Mo	0.1	200	5	0.01	scale	poly	True

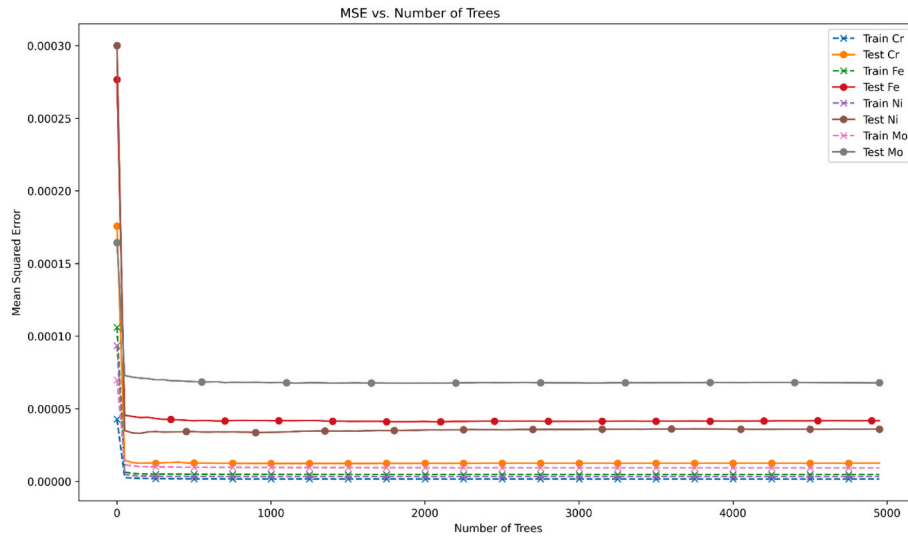


Fig. 4. MSE as a function of the number of trees in random forest.

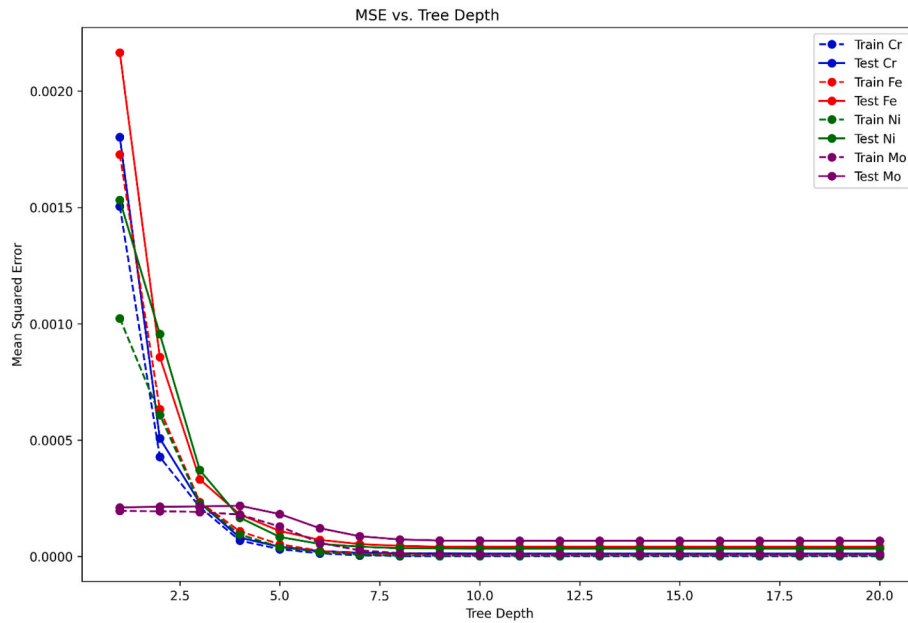


Fig. 5. MSE as a function of tree depth in random forest.

Table 8

Optimal hyperparameters for artificial neural network.

Hyperparameter	Optimal Value
Hidden Units	1024
Dropout Rate	0.1
Learning Rate	0.001
Weight Decay	0.00001
Epochs	250
Optimizer	adam

internal validation indicator, the RF model still showed overfitting for Mo. This limitation is likely associated with the relatively low and narrow concentration range of Mo in the training dataset, which may make the model more sensitive to small variations in the normalized Mo response. Random Forest models are generally designed to reduce variance by averaging multiple decision trees, and increasing the number of trees alone is not expected to cause progressive overfitting because the generalization error tends to converge as the forest size

increases [45]. However, overfitting may still occur when a weak or low-concentration output variable is learned together with stronger output variables under a single global optimization criterion. Therefore, future implementations should refine the training strategy by using element-wise or weighted validation criteria and by increasing the density of simulated samples in the low-Mo region. In contrast, the ANN model sustains an R^2 value of 0.99 in both training and testing phases. When the test dataset is evaluated in terms of MSE scores, SVR and ANN models exhibit similar performance to their training results. However, the RF model demonstrates a significant increase in MSE, reflecting a decline in its predictive performance.

Finally, all trained models were tested and compared using experimentally obtained spectra of real stainless steels. This comparison was not conducted through performance metrics but directly by comparing the experimentally determined elemental distributions of stainless steels with the predictions made by the models. Table 11 compares the elemental compositions predicted by the machine learning models from experimental ED-XRF spectra with the reference compositions of the stainless-steel samples. Here, the reference compositions correspond to

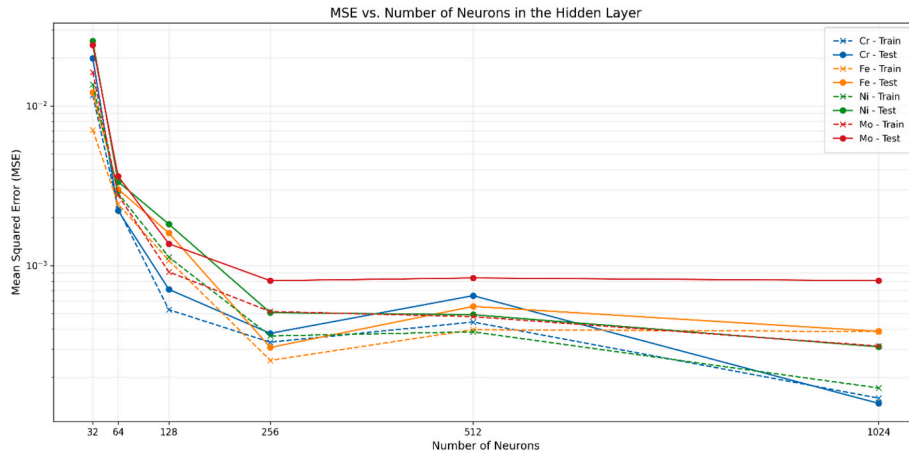


Fig. 6. MSE as a function of the number of neurons in ANN.

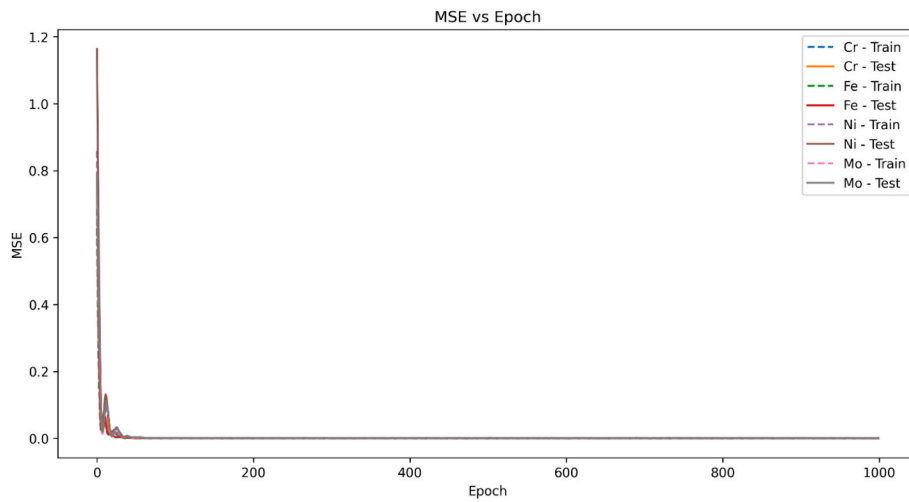


Fig. 7. MSE as a function of epochs with the optimal number of neurons in ANN.

Table 9

Performance metrics of models on training dataset.

Model	Element	MSE	R ²
SVR	Cr	2.0393×10^{-5}	0.9899
	Fe	2.3508×10^{-5}	0.9960
	Ni	2.1427×10^{-5}	0.9940
	Mo	3.4464×10^{-5}	0.8256
RF	Cr	1.6946×10^{-6}	0.9992
	Fe	4.8489×10^{-6}	0.9992
	Ni	3.2986×10^{-6}	0.9991
	Mo	9.6235×10^{-6}	0.9513
ANN	Cr	1.4798×10^{-5}	0.9999
	Fe	3.5812×10^{-5}	0.9996
	Ni	1.4982×10^{-5}	0.9999
	Mo	3.3415×10^{-5}	0.9997

Table 10

Performance metrics of models on test dataset.

Model	Element	MSE	R ²
SVR	Cr	2.5717×10^{-5}	0.9887
	Fe	2.3574×10^{-5}	0.9966
	Ni	2.1760×10^{-5}	0.9955
	Mo	3.3235×10^{-5}	0.8290
RF	Cr	1.2391×10^{-5}	0.9946
	Fe	4.1763×10^{-5}	0.9941
	Ni	3.3897×10^{-5}	0.9931
	Mo	6.8036×10^{-5}	0.6500
ANN	Cr	1.5325×10^{-4}	0.9999
	Fe	3.6881×10^{-5}	0.9997
	Ni	2.2423×10^{-5}	0.9998
	Mo	5.8995×10^{-5}	0.9994

the certified composition for SS316 and to WD-XRF/SQX results for the remaining samples. For predictions based on experimental spectra from real stainless steels, all algorithms provide comparable results for Cr, Fe, and Ni elements. However, for Mo, which is present at low concentrations in these stainless steels, the ANN model shows the closest agreement with the reference values. Finally, considering both the training and testing scores as well as the validations performed with experimental data, these results indicate that the ANN model outperforms the other two models.

The calculated uncertainty ranges support the stability of the machine learning predictions under the adopted uncertainty estimation

procedures. For the ANN models, uncertainty was determined using the Monte Carlo Dropout method [52]. Although dropout layers are typically used to prevent overfitting during training, they remained active during the testing phase for this work. Each spectrum was processed 100 times with stochastic dropout activated, and the standard deviation of these iterations was recorded as the prediction uncertainty. The RF is an ensemble algorithm consisting of many independent decision trees. Prediction uncertainty was calculated from the variance among the individual predictions produced by 1000 trees [45,53]. Higher consensus among the trees results in lower uncertainty, whereas samples that lead to varying results across the trees show higher uncertainty levels. For the

Table 11

Comparison of reference compositions and machine learning predictions obtained from experimental ED-XRF spectra.

Steel type	Element	Reference composition (%)	SVR	RF	ANN
SS316	Cr	18.45 ± 0.05	14.92 ± 0.06	14.09 ± 0.99	14.53 ± 0.18
	Fe	65.19 ± 0.07	66.75 ± 0.14	67.14 ± 1.19	67.18 ± 0.2
	Ni	12.18 ± 0.05	13.53 ± 0.04	14.71 ± 1.42	14.31 ± 0.23
	Mo	2.38 ± 0.01	1.48 ± 0.04	1.23 ± 1.15	1.18 ± 0.05
SS303	Cr	18.9 ± 0.009	16.51 ± 0.11	16 ± 0.96	16.18 ± 0.17
	Fe	68.4 ± 0.014	70.97 ± 0.1	71.24 ± 1.24	71.68 ± 0.3
	Ni	8.32 ± 0.008	8.42 ± 0.08	9.12 ± 1.23	9.10 ± 0.21
	Mo	0.46 ± 0.003	0.86 ± 0.05	0.81 ± 1.12	0.25 ± 0.07
SS304	Cr	18.9 ± 0.012	16.61 ± 0.12	16.04 ± 0.89	16.34 ± 0.18
	Fe	70.6 ± 0.020	71.48 ± 0.1	71.43 ± 1.29	72.27 ± 0.31
	Ni	7.74 ± 0.007	8 ± 0.09	8.82 ± 1.27	8.64 ± 0.22
	Mo	0.32 ± 0.002	0.66 ± 0.05	0.88 ± 1.12	0.43 ± 0.07
SS430	Cr	17.4 ± 0.009	13.69 ± 0.11	13.81 ± 0.92	14.07 ± 0.34
	Fe	81.9 ± 0.021	82.37 ± 0.11	82.43 ± 1.24	83.02 ± 0.66
	Ni	0.125 ± 0.005	0.96 ± 0.09	0.66 ± 1.07	0.35 ± 0.36
	Mo	<LOD	1 ± 0.05	0.29 ± 0.76	0 ± 0.04

SVR models, which have a deterministic structure, a bootstrapping approach was preferred [54–56]. Thirty different SVR models were trained using random subsets of the training data selected with replacement. The stability of this ensemble for the same input was evaluated, and the standard deviation of the predictions was defined as the uncertainty margin.

Table 11 allows a direct comparison of prediction reliability across the three models. SVR produced the narrowest intervals, with a mean uncertainty of ± 0.08 wt%, attributable to the consistency of the bootstrapped ensemble across resampled training subsets. RF performed poorly in this regard, averaging ± 1.12 wt% across all elements, well above both alternatives. For Mo, RF uncertainties ranged from ± 0.76 to ± 1.15 wt%; relative to the predicted concentrations, these uncertainty margins correspond to approximately 94–138% of the predicted value, effectively rendering the RF predictions for Mo statistically indistinguishable from zero. The ANN yielded a mean of ± 0.22 wt%. Although wider than SVR, its Mo uncertainty stayed below ± 0.07 wt% across all four samples, which no other model achieved alongside comparable prediction accuracy.

However, some deviations between predicted and experimental results exist due to several factors. First, simulation uncertainty in MCNP is governed by the total particle count. In this work, the use of 10^9 particles kept the simulation statistical uncertainty below 0.1% for every element. Second, the uncertainty of the experimental XRF measurement was reduced by taking the average of five spectra for each sample, with each measurement lasting 300 s. The resulting average uncertainty is below 0.3%. Despite these efforts, among the main causes of deviation are the differences between the simulated and experimental background levels and the presence of characteristic peaks from the gold target of the X-ray tube in the experimental spectrum, which were not captured in the simulated source spectrum, as shown in Fig. 8. Simulated spectra generally include only physical interactions and detector broadening

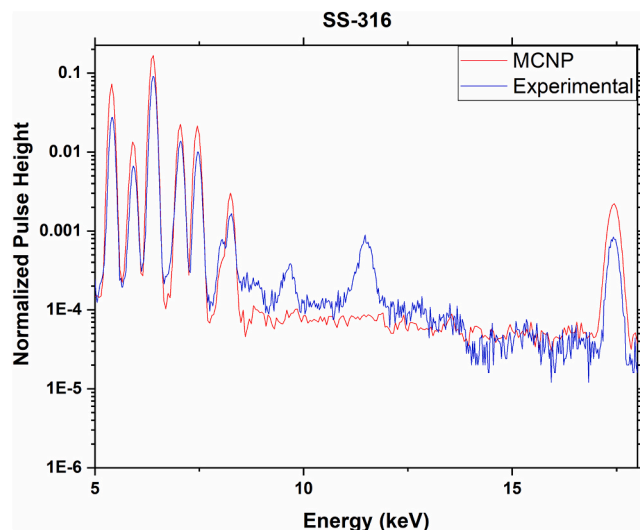


Fig. 8. Comparison of experimental and MCNP-simulated ED-XRF spectra for SS316 stainless steel.

explicitly implemented in the model, whereas experimental spectra may contain additional detector-response artifacts such as electronic noise, incomplete charge collection, pulse pile-up, escape peaks, and sum peaks. Therefore, residual discrepancies between MCNP simulations and measured spectra can arise from unmodelled detector response effects as well as from uncertainties in the source spectrum modelling and cross-section data [57–59].

4. Conclusion

This study aimed to determine the elemental distribution of stainless steel with machine learning algorithms. Machine learning models developed for this purpose were trained using spectra generated by MCNP simulations of 400 stainless steel combinations. To our knowledge, this study is among the first to compare machine learning algorithms trained with MCNP-simulated XRF spectra for stainless-steel analysis. To obtain stainless steel spectra through MCNP simulations, a simulation geometry consistent with the experimental setup was designed.

Peak normalization was applied to the simulated XRF spectra, and normalized characteristic peak areas were calculated. Datasets were created with normalized characteristic peak areas as independent variables and the percentage weights of the elements as dependent variables. Using these datasets, three different machine learning models were trained. The trained models were validated using experimentally obtained spectra of real stainless steels. Among the models, ANN showed the closest agreement with both simulation-derived and experimental datasets.

The results of this study suggest that models trained with XRF spectra can be developed not only for stainless steel but also for other materials. Furthermore, it was observed that using only simulation data to create datasets for machine learning models can yield accurate results for practical applications. This shows that it is possible to train usable machine learning models with simulation data alone. Integrating the trained machine learning models into portable or desktop XRF systems could enable devices capable of performing qualitative and quantitative analysis with a single experimental spectrum, eliminating the need for repeated spectrum acquisition. Despite these advantages, machine learning algorithms trained with this method are likely to produce inaccurate predictions when applied to elemental concentrations outside the range of the training dataset. For example, since the Mo range in our dataset is 0–4%, a sample containing 10% Mo is likely to yield inaccurate predictions.

CRediT authorship contribution statement

Kaan Yavaş: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation. **Ulaş Keskin:** Writing – original draft, Methodology, Investigation. **Ozan Toker:** Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation. **Özgür Akçalı:** Supervision, Methodology, Investigation. **H. Birtan Kavanoz:** Supervision, Methodology, Methodology. **Orhan İçelli:** Writing – review & editing, Writing – original draft, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This study was funded by the Scientific and Technological Research Council of Türkiye (TUBITAK) 1002-A Grant No 123F246.

We are grateful to the Scientific Research Project Center at Yildiz Technical University for supporting this research under the Project No. FBA-2024-5902.

Data availability

Data will be made available on request.

References

- N.R. Baddoo, Stainless steel in construction: a review of research, applications, challenges and opportunities, *J. Constr. Steel Res.* 64 (11) (2008) 1199–1206, <https://doi.org/10.1016/j.jcsr.2008.07.011>.
- M.M. Mourad, H.A. Saudi, M.M. Eissa, M.Y. Hassaan, M.A. Abdel-Latif, Modified austenitic stainless-steel alloys for shielding nuclear reactors, *Prog. Nucl. Energy* 142 (2021), <https://doi.org/10.1016/j.pnucene.2021.104009>.
- Panagopoulos, A., Loizidou, M., & Haralambous, K. J. (2020). Stainless Steel in Thermal Desalination and Brine Treatment: Current Status and Prospects. In *Metals and Materials International* (Vol. 26, Number 10, pp. 1463–1482). Korean Institute of Metals and Materials. Doi:10.1007/s12540-019-00398-w.
- B. Rossi, Discussion on the use of stainless steel in constructions in view of sustainability, *Thin-Walled Struct.* 83 (2014) 182–189, <https://doi.org/10.1016/j.tws.2014.01.021>.
- A. Saha Podder, A. Bhanja, Applications of stainless steel in automobile industry, *Adv. Mat. Res.* 794 (2013) 731–740, <https://doi.org/10.4028/www.scientific.net/AMR.794.731>.
- M. Kemp, A. van Bennekom, F.P.A. Robinson, Evaluation of the corrosion and mechanical properties of a range of experimental Cr–Mn stainless steels, *Mater. Sci. Eng. A* 199 (2) (1995) 183–194, [https://doi.org/10.1016/0921-5093\(94\)09694-5](https://doi.org/10.1016/0921-5093(94)09694-5).
- H.H. Pariser, N.R. Backeberg, O.C.M. Masson, J.C.M. Bedder, Changing nickel and chromium stainless steel markets – a review, *J. South Afr. Inst. Min. Metall.* 118 (6) (2018) 563–568, <https://doi.org/10.17159/2411-9717/2018/V118N6A1>.
- E. Calparsoro, M. Maguregui, H. Morillas, G. Arana, J.G. Iñáñez, Non-destructive screening methodology based on ED-XRF for the classification of medieval and post-medieval archaeological ceramics, *Ceram. Int.* 45 (8) (2019) 10672–10683, <https://doi.org/10.1016/j.ceramint.2019.02.138>.
- M.L. Carvalho, S. Santiago, M.L. Nunes, Assessment of the essential element and heavy metal content of edible fish muscle, *Anal. Bioanal. Chem.* 382 (2) (2005) 426–432, <https://doi.org/10.1007/s00216-004-3005-3>.
- M. Warlo, G. Bark, C. Wanhainen, I. McElroy, A. Björling, U. Johansson, Extreme-resolution synchrotron X-ray fluorescence mapping of ore samples, *Ore Geol. Rev.* 140 (2022), <https://doi.org/10.1016/j.oregeorev.2021.104620>.
- R.M. Rousseau, Corrections for matrix effects in X-ray fluorescence analysis—A tutorial, *Spectrochim. Acta B At. Spectrosc.* 61 (7) (2006) 759–777, <https://doi.org/10.1016/J.SAB.2006.06.014>.
- Sitko, R., & Zawisz, B. (2012). Quantification in X-Ray Fluorescence Spectrometry. In *X-Ray Spectroscopy*. InTech. Doi:10.5772/29367.
- I. Sørensen, A calculated background correction method for XRF, *Spectrochim. Acta B At. Spectrosc.* 36 (3) (1981) 243–249, [https://doi.org/10.1016/0584-8547\(81\)80026-2](https://doi.org/10.1016/0584-8547(81)80026-2).
- S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytrcek, D. Zschesche, Geant4—a simulation toolkit, *Nucl. Instrum. Methods Phys. Res., Sect. A* 506 (3) (2003) 250–303, [https://doi.org/10.1016/S0168-9002\(03\)01368-8](https://doi.org/10.1016/S0168-9002(03)01368-8).
- C. Ahdida, D. Bozzato, D. Calzolari, F. Cerutti, N. Charitonidis, A. Cimmino, A. Coronetti, G.L. D'Alessandro, A. Donadon Servelle, L.S. Esposito, R. Froeschl, R. García Alfía, A. Gerbershagen, S. Gilardoni, D. Horváth, G. Hugo, A. Infantino, V. Kouskoura, A. Lechner, M. Witorski, New capabilities of the FLUKA multi-purpose code, *Front. Phys.* 9–2021 (2022), <https://doi.org/10.3389/fphy.2021.788253>.
- G. Battistoni, T. Boehlen, F. Cerutti, P.W. Chin, L.S. Esposito, A. Fassò, A. Ferrari, A. Lechner, A. Empl, A. Mairani, A. Mereghetti, P.G. Ortega, J. Ranft, S. Roesler, P. R. Sala, V. Vlachoudis, G. Smirnov, Overview of the FLUKA code, *Ann. Nucl. Energy* 82 (2015), <https://doi.org/10.1016/j.anucene.2014.11.007>.
- Werner, C. J., Armstrong, J. C., Brown, F. B., Bull, J. S., Casswell, L., Cox, L. J., Dixon, D. A., Forster III, R. A., Goorley, J. T., Hughes III, H. G., Favorite, J. A., Martz, R. L., Mashnik, S. G., Rising, M. E., Solomon Jr., C. J., Sood, A., Sweezy, J. E., Zukaitis, A. J., Anderson, C. A., ... Kiedrowski, B. C. (2017). MCNP User's Manual Code Version 6.2 (Number LA-UR-17-29981). <https://permlink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-17-29981>.
- S. An, S. Reza, B. Norlin, C. Fröjd, G. Thungström, Signal-to-noise ratio optimization in X-ray fluorescence spectrometry for chromium contamination analysis, *Talanta* 230 (2021), <https://doi.org/10.1016/j.talanta.2021.122236>.
- T. Trojek, T. Čechák, Use of MCNP code in energy dispersive X-ray fluorescence, *Nucl. Instrum. Methods Phys. Res., Sect. B* 263 (1 SPEC. ISS.) (2007) 72–75, <https://doi.org/10.1016/j.nimb.2007.04.063>.
- L. Vincze, K. Janssens, F. Adams, M.L. Rivers, K.W. Jones, A general Monte Carlo simulation of ED-XRF spectrometers. II: Polarized monochromatic radiation, homogeneous samples, *Spectrochim. Acta B At. Spectrosc.* 50 (2) (1995) 127–147, [https://doi.org/10.1016/0584-8547\(94\)00124-E](https://doi.org/10.1016/0584-8547(94)00124-E).
- M. Wang, Y. Gu, H. Lu, L. Ge, Q. Zhang, G. Zeng, Matrix effect correction method based on the main spectral parameters for rock samples in an in situ energy dispersive X-ray fluorescence analysis, *Spectrochim. Acta B At. Spectrosc.* 193 (2022) 106438, <https://doi.org/10.1016/J.SAB.2022.106438>.
- T. Schoonjans, V.A. Solé, L. Vincze, M. Sanchez Del Rio, K. Appel, C. Ferrero, A general Monte Carlo simulation of energy-dispersive X-ray fluorescence spectrometers - Part 6. Quantification through iterative simulations, *Spectrochim. Acta B At. Spectrosc.* 82 (2013) 36–41, <https://doi.org/10.1016/j.sab.2012.12.011>.
- I. Szalóki, A. Gerényi, G. Radócz, Reverse Monte Carlo iterative algorithm for quantification of X-ray fluorescence analysis based on MCNP6 simulation code, *X-Ray Spectrom.* 49 (5) (2020) 587–595, <https://doi.org/10.1002/xrs.3154>.
- T. Trojek, Iterative Monte Carlo procedure for quantitative X-ray fluorescence analysis of copper alloys with a covering layer, *Radiat. Phys. Chem.* 167 (2020), <https://doi.org/10.1016/J.RADPHYSCH.2019.04.044>.
- K. Yavaş, O. Toker, Ö. Akcalı, O. İçelli, NXSIM: a normalized XRF standardless iterative simulation method for qualitative and quantitative XRF analysis via MCNP, *Nucl. Instrum. Methods Phys. Res., Sect. B* 565 (2025) 165739, <https://doi.org/10.1016/J.NIMB.2025.165739>.
- B. Bilmeze, O. Toker, S. Alp, E. Öz, O. İçelli, A comparative study on applicability and efficiency of machine learning algorithms for modeling gamma-ray shielding behaviors, *Nucl. Eng. Technol.* 54 (1) (2022) 310–317, <https://doi.org/10.1016/j.net.2021.07.031>.
- Gogas, P., & Papadimitriou, T. (2021). Machine Learning in Economics and Finance. In *Computational Economics* (Vol. 57, Number 1). Springer. Doi:10.1007/s10614-021-10094-w.
- A. Bos, M. Bos, W.E. van der Linden, Artificial neural networks as a multivariate calibration tool: modeling the Fe–Cr–Ni system in x-ray fluorescence spectroscopy, *Anal. Chim. Acta* 277 (2) (1993) 289–295, [https://doi.org/10.1016/0003-2670\(93\)80441-M](https://doi.org/10.1016/0003-2670(93)80441-M).
- I. Allegretta, B. Marangoni, P. Manzari, C. Porfido, R. Terzano, O. De Pascale, G. S. Senesi, Macro-classification of meteorites by portable energy dispersive X-ray fluorescence spectroscopy (pED-XRF), principal component analysis (PCA) and machine learning algorithms, *Talanta* 212 (2020), <https://doi.org/10.1016/j.talanta.2020.120785>.
- C. Jones, N.S. Daly, C. Higgitt, M.R.D. Rodrigues, Neural network-based classification of X-ray fluorescence spectra of artists' pigments: an approach leveraging a synthetic dataset created using the fundamental parameters method, *Heritage Science* 10 (1) (2022) 1–14, <https://doi.org/10.1186/S40494-022-00716-3>.
- A. Rakotondrajoa, M. Radtke, Machine learning based quantification of synchrotron radiation-induced x-ray fluorescence measurements—a case study, *Mach. Learn.: Sci. Technol.* 2 (2) (2021) 025004, <https://doi.org/10.1088/2632-2153/abc9fb>.
- A.S. Lee, W.S. Chao, S.Y.H. Liou, R. Tiedemann, B. Zolitschka, L. Lembke-Jene, Quantifying calcium carbonate and organic carbon content in marine sediments from XRF-scanning spectra with a machine learning approach, *Sci. Rep.* 12 (1) (2022), <https://doi.org/10.1038/s41598-022-25377-x>.
- Zhao, W., Ai, X., & Zhao, H. (2023). Quantitative analysis of energy-dispersive X-ray fluorescence spectroscopy based on machine learning and a generative data enhancement technique. *Applied Optics*, Vol. 62, Issue 36, Pp. 9476–9485, 62(36), 9476–9485. Doi:10.1364/AO.506027.
- Cook, M. T. (2015). Hybrid K-edge Densitometry as a Method for Materials Accountancy Measurements in Pyrochemical Reprocessing. Doctoral Dissertations. https://trace.tennessee.edu/utk_graddiss/3329.
- A. Poškus, Evaluation of computational models and cross sections used by MCNP6 for simulation of characteristic X-ray emission from thick targets bombarded by kiloelectronvolt electrons, *Nucl. Instrum. Methods Phys. Res., Sect. B* 383 (2016), <https://doi.org/10.1016/j.nimb.2016.06.017>.

- [36] S. Jung, W. Sung, J. Lee, S.J. Ye, MCNP6.1 simulations for low-energy atomic relaxation: Code-to-code comparison with GATEv7.2, PENELOPE2014, and EGSnrc, Nucl. Instrum. Meth. Phys. Res., Sect. B (2018) 415, <https://doi.org/10.1016/j.nimb.2017.11.024>.
- [37] Parsons, D. K. (2022). Some Notes on EPRDATA and MCNP (No. LA-UR-22-27131). Los Alamos National Laboratory (LANL), Los Alamos, NM (United States). Doi: 10.2172/1877996.
- [38] S. Jung, Monte Carlo modeling of gold nanoparticles detection limits of benchtop three-dimensional L- and K-shell X-ray fluorescence mapping systems, X-Ray Spectrom. 52 (1) (2023), <https://doi.org/10.1002/xrs.3302>.
- [39] Hughes III, H. G. (2015). An Electron/Photon/Relaxation Data Library for MCNP6. Los Alamos National Laboratory.
- [40] A. Gupta, A. Sharma, A. Goel, Review of regression analysis models, Int. J. Eng. Res. V6 (08) (2017), <https://doi.org/10.17577/IJERTV6IS080060>.
- [41] G. Van Rossum, F.L. Drake, Python 3 Reference Manual, CreateSpace (2009).
- [42] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Support vector regression machines, in: M.C. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems, vol. 9, MIT Press, 1996.
- [43] K. Were, D.T. Bui, Ø.B. Dick, B.R. Singh, A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape, Ecol. Ind. 52 (2015) 394–403, <https://doi.org/10.1016/j.ecolind.2014.12.028>.
- [44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830.
- [45] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32, <https://doi.org/10.1023/A:1010933404324>.
- [46] P. Gulati, A. Sharma, M. Gupta, Theoretical study of decision tree algorithms to identify pivotal factors for performance improvement: a review, Int. J. Comput. Appl. 141 (14) (2016) 19–25, <https://doi.org/10.5120/ijca2016090926>.
- [47] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines, Ore Geol. Rev. 71 (2015) 804–818, <https://doi.org/10.1016/j.oregeorev.2015.01.001>.
- [48] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (6088) (1986) 533–536, <https://doi.org/10.1038/323533a0>.
- [49] Z.S. Kadhim, H.S. Abdullah, K.I. Ghatwan, Artificial neural network hyperparameters optimization: a survey, Int. J. Online Biomed. Eng. 18 (15) (2022) 59–87, <https://doi.org/10.3991/ijoe.v18i15.34399>.
- [50] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: an imperative style, high-performance deep learning library. In Proceedings of the 33rd International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc. Doi:10.48550/arXiv.1912.01703.
- [51] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) 1–24, <https://doi.org/10.7717/peerj-cs.623>.
- [52] Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning (pp. 1050–1059). PMLR. <https://proceedings.mlr.press/v48/gal16.html>.
- [53] N. Meinshausen, Quantile Regression Forests, J. Mach. Learn. Res. 7 (2006) 983–999.
- [54] W.Y. Choi, D.H. Choi, K.J. Cha, Robust estimation of support vector regression via residual bootstrap adoption, J. Mech. Sci. Technol. 29 (1) (2015) 279–289, <https://doi.org/10.1007/s12206-014-1234-8>.
- [55] S. Lee, A. Ahmad, G. Jeon, Combining bootstrap aggregation with support vector regression for small blood pressure measurement, J. Med. Syst. 42 (4) (2018), <https://doi.org/10.1007/s10916-018-0913-x>.
- [56] I.D. Lins, E.L. Droguett, M.D.C. Moura, E. Zio, C.M. Jacinto, Computing confidence and prediction intervals of industrial equipment degradation by bootstrapped support vector regression, Reliab. Eng. Syst. Saf. 137 (2015) 120–128, <https://doi.org/10.1016/j.ress.2015.01.007>.
- [57] X. Llovet, L. Sorbier, C.S. Campos, E. Acosta, F. Salvat, Monte Carlo simulation of x-ray spectra generated by kilo-electron-volt electrons, J. Appl. Phys. 93 (7) (2003) 3844–3851, <https://doi.org/10.1063/1.1545154>.
- [58] F. Scholze, M. Procop, Modelling the response function of energy dispersive X-ray spectrometers with silicon detectors, X-Ray Spectrom. 38 (4) (2009) 312–321, <https://doi.org/10.1002/XRS.1165>.
- [59] R. Taleei, M. Shahriari, Monte Carlo simulation of X-ray spectra and evaluation of filter effect using MCNP4C and FLUKA code, Appl. Radiat. Isot. 67 (2) (2009) 266–271, <https://doi.org/10.1016/J.APRADISO.2008.10.007>.