Journal of Intelligent Decision and Computational Modelling

www.jidcm.reapress.com

J. Intell. Decis. Comput. Model. Vol. 2, No. 2 (2026) 155–164.

Paper Type: Original Article

Identification of Porphyry Copper Anomalies Using Support Vector Regression

Farzad Moradpouri[1,*], Mohammad Bagher Dowlatshahi[2]

[1] Department of Mining Engineering, Faculty of Engineering, Lorestan University; moradpouri.fa@lu.ac.ir.

[2] Department of Computer Engineering, Faculty of Engineering, Lorestan University; dowlatshahi.mb@lu.ac.ir.

Citation:

Received: 01 October 2025 *Revised: 22 December 2025* *Accepted: 18 March 2026*	Moradpouri, F., & Dowlatshahi, M. B. (2026). Identification of porphyry copper anomalies using support vector regression. *Journal of intelligent decision and computational modelling*, 2(2), 155-164.

Abstract

In the present study, data obtained from 83 rock samples collected from a porphyry copper mineralization zone in Canada were analyzed. First, the F-test feature selection method was employed to identify and determine the significant features, whereby five elements, gold (Au), molybdenum (Mo), silver (Ag), iron (Fe), and lead (Pb), were selected as the most important features for implementing the regression models. Subsequently, 70% of the data were allocated for training, and the remaining data were used for testing the models. To predict copper values, several methods, including the linear regression model, decision tree regression model, bagging ensemble tree model, boosting ensemble tree model, and a three-layer neural network model, were applied. The data were then predicted using the novel Projection Support Vector Regression (PSVR) model, and the Root Mean Square Error (RMSE), R-squared (R^2), and training time of each method were analyzed. Furthermore, we generated maps of prospective mineralized zones in a Geographic Information System (GIS) environment based on the predicted copper values from the aforementioned methods. Finally, we compared the results of all methods, demonstrating the efficiency and superiority of the proposed PSVR method over the other investigated approaches.

Keywords: Porphyry copper, Geochemical prospective zones, Support vector regression, Geographic information system.

1 | Introduction

Identifying and mapping geochemical anomaly prospective zones is among the most important stages of geochemical exploration studies [1]. Preparing such maps is essential for decision-making regarding the subsequent stages of an exploration project. On the other hand, during sampling from different geochemical environments (rock, soil, stream sediments, etc.), a large number of samples are usually collected. The analytical results of these samples form a relatively large matrix consisting of numerous samples and various variables, whose analysis requires accurate processing and interpretation [2].

Researchers have employed various methods for interpreting geochemical data, identifying prospective zones, and mapping them. Machine Learning (ML) methods are among the common approaches used for pattern

Corresponding Author: moradpouri.fa@lu.ac.ir

https://doi.org/10.48314/jidcm.v2i2.92

recognition in data and identification of geochemical anomalies [3]. Numerous researchers have applied different multivariate analysis and ML techniques in their studies for data analysis and preparation of geochemical anomaly maps [4]. Neural network-based methods, Support Vector Machine (SVM) techniques, and multi-criteria decision-making methods are among these approaches [5].

This study aims to predict copper concentrations in a porphyry copper mineralization zone located in British Columbia, Canada, using data obtained from rock samples. For this purpose, we first selected significant variables using the F-test feature selection method. Subsequently, copper values were predicted using six different regression models, and after comparing the results of the different methods, prospective zone maps were generated and presented.

1.1 | Location of the Study Area

The location of the study area is shown in *Fig. 1*, which is situated in the Stikine River region in northwestern British Columbia, Canada. The potential mineralization is of the porphyry deposit type and is associated with magmatic and hydrothermal activities occurring at depth and near the Earth's surface. Porphyry deposits may contain millions of tons of metals such as copper, gold, and molybdenum, with grades ranging from ppm levels to 2%, making them economically significant for countries with such resources.

The samples consisted of 83 representative rock samples collected using the grab sampling method from existing outcrops, particularly those containing sulfide minerals such as pyrite and other metal-bearing sulfides. The samples were analyzed using the ICP-MS method and included 36 elements: Au, Cu, Zn, Mo, Pb, Co, Mn, Ag, Ni, Th, Sr, Fe, As, Bi, V, Cd, Sb, La, Cr, Ca, P, Ti, B, Mg, Ba, K, W, Al, Na, S, Sc, Hg, Tl, Te, Ga, and Se.

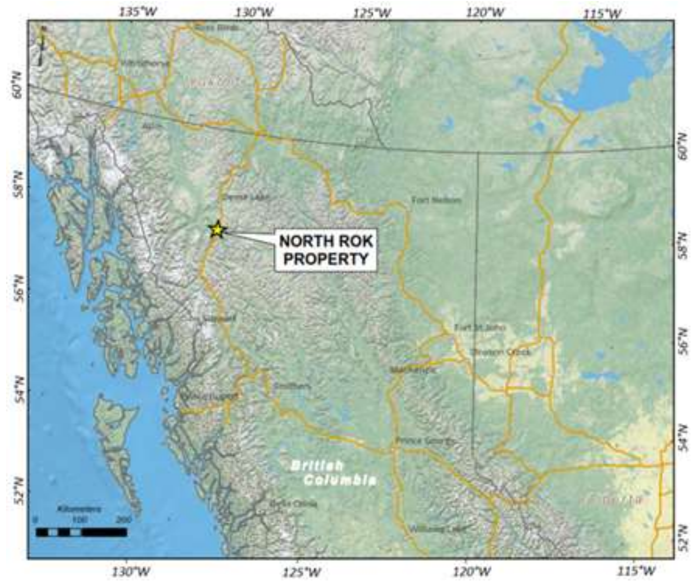


Fig. 1. Location of the study area.

2 | Support Vector Regression

SVMs are supervised learning algorithms proposed by Cortes and Vapnik in 1995 [6]. This section presents an overview of SVMs for continuous target variables, known as epsilon-based Support Vector Regression (ε-SVR).

This section presents an overview of SVMs for continuous target variables, called epsilon-based SVRs (ε-SVRs). SVR attempts to approximate a function f_x that can handle as many examples from the training set as possible, with a maximum deviation ε from the target, while remaining as smooth as possible. Suppose we model an output target y as a function of n input variables x and have a training set of length N, for example, $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$ where $x_i \in \mathbb{R}^n$ and $y_i \in \mathbb{R}$ for $i = 1,2, \ldots, N$. In this case, SVR approximates the relationship between the input and target variables using *Eq. 1*.

$$f(x)=\langle w.\Phi(x)\rangle+b\ , \quad (1)$$

where ⟨·,·⟩ denotes the inner product, w is the weight vector, b is the bias term, and $\Phi(x)$ is the kernel function for mapping samples from the input space to a higher-dimensional feature space. In fact, the objective is to minimize the following optimization problem (*Eq. 2*):

$$\min\frac{\|w^2\|}{2}$$
$$\text{subject to}\ \begin{cases} y_i-\langle w.\Phi(x_i)\rangle-b\le\varepsilon\ ,\\ \langle w.\Phi(x_i)\rangle+b-y_i\le\varepsilon\end{cases} \quad (2)$$

where ε is the allowable error, and $\| w \|$ describes the Euclidean norm of the weight vector. The above optimization problem is feasible only when there exists a function f(x) that can approximate all training data with an accuracy of ε.

2.1 | Proposed Support Vector Regression Modeling Method

Fig. 2 illustrates the steps of the proposed SVR modeling procedure for predicting the target variable. The procedure begins by setting different SVR model parameters. Next, the dataset containing the dependent and independent variables is prepared, and the SVR model is trained on the training dataset and then evaluated on the test data. The SVR model performance is then analyzed using statistical criteria.

Finally, the prediction model proposes the final conditions for solving the problem.

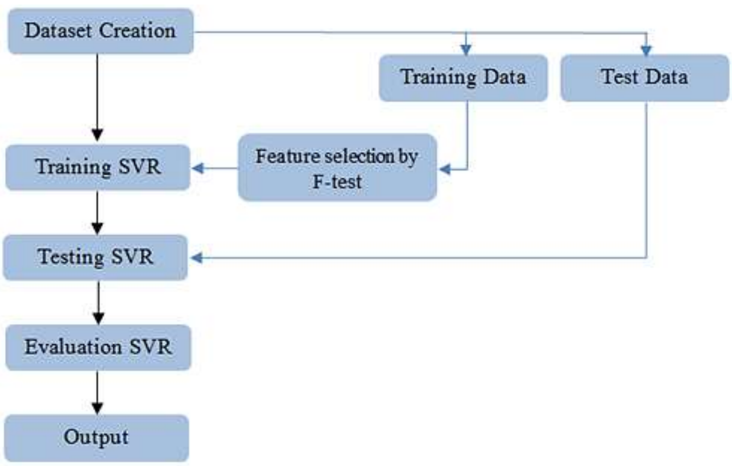


Fig. 2. Flowchart of the proposed SVR modeling procedure.

2.2 | Feature Selection

All researchers working in the fields of Artificial Intelligence (AI) and ML face the problem of selecting important features within a dataset and removing or ignoring insignificant features (which do not contribute substantially to the accurate prediction of target values for new samples and consequently to the performance of learning models). Feature selection is a key concept in ML and plays an important role in the optimal performance of learning models. The F-test is a statistical test used to compare the variances of two or more populations. It employs the F distribution, which is a probability distribution, to test the hypothesis that two variances are equal. In fact, the F-test determines whether there is a significant difference between the variances of two or more populations. To perform the F-test, calculate the F statistic by dividing the variance of one group by the variance of another group [7]. The resulting value is then compared with the F distribution to determine whether there is a significant difference between the variances of the two groups. Table 1 presents the preliminary statistical analysis of the rock samples in the present study. The F-test feature selection algorithm was applied to a dataset of 83 rock samples and 36 variables (features), and the results are shown in Fig. 3. As shown, Au, Mo, Ag, Fe, and Pb are the five most significant features among all variables. Therefore, these five variables were used as the most important predictor features for predicting copper concentrations using the regression models. As shown in *Fig. 3*, gold and silver, identified as important variables associated with copper based on geological evidence and soil data modeling, are also among the five variables selected by the F-test algorithm.

Table 1. Basic statistical parameters of the 83 rock samples used in this study.

Variable	Minimum	Maximum	Mean	Standard Deviation	Variable	Minimum	Maximum	Mean	Standard Deviation
Cu (ppm)	6.6	1982	79.93	134.07	P (ppm)	10	10120	1054.78	849.46
Mo (ppm)	0.4	95.2	3.07	4.39	La (ppm)	1	149	9.08	7.49
Pb (ppm)	0.5	1117	20.46	36.13	Cr (ppm)	3	87	29.91	10.77
Zn (ppm)	6	1795	109.87	103.40	Mg (ppm)	1000	20400	7464.34	2967.93
Ag (ppm)	0.1	5	0.37	0.36	Ba (ppm)	19	1119	197.28	118.55
Ni (ppm)	2.2	175.2	32.38	18.21	Ti (ppm)	30	6880	1166.03	645.64
Co (ppm)	1.1	320	15.71	12.64	B (ppm)	1	39	18.01	5.79
Mn (ppm)	30	10000	837.37	883.26	Al (ppm)	800	56800	20438.44	6844.29
Fe (ppm)	3400	207800	47396.81	20394.86	Na (ppm)	40	6620	168.76	234.65
As (ppm)	0.5	1033	19.49	44.35	K (ppm)	10	5200	827.33	423.6
Au (ppm)	0.001	0.936	0.012	0.042	W (ppm)	0.1	1.5	0.19	0.13
Th (ppm)	0.1	6.7	1.13	0.79	Hg (ppm)	0.01	0.93	0.057	0.07
Sr (ppm)	5	524	46	42.16	Sc (ppm)	0.4	26.8	4.55	2.41
Cd (ppm)	0.1	46.2	0.994	2.32	Tl (ppm)	0.1	1.1	0.10	0.03
Sb (ppm)	0.1	2.5	0.49	0.28	S (ppm)	500	20600	939.27	1560.05
Bi (ppm)	0.1	15.6	0.46	0.77	Ga (ppm)	1	27	9.54	3.53
V (ppm)	3	302	89.12	31.05	Se (ppm)	0.5	36.8	1.24	2.22
Ca (ppm)	200	59300	4720.64	5633.94	Te (ppm)	0.2	6.8	0.28	0.34

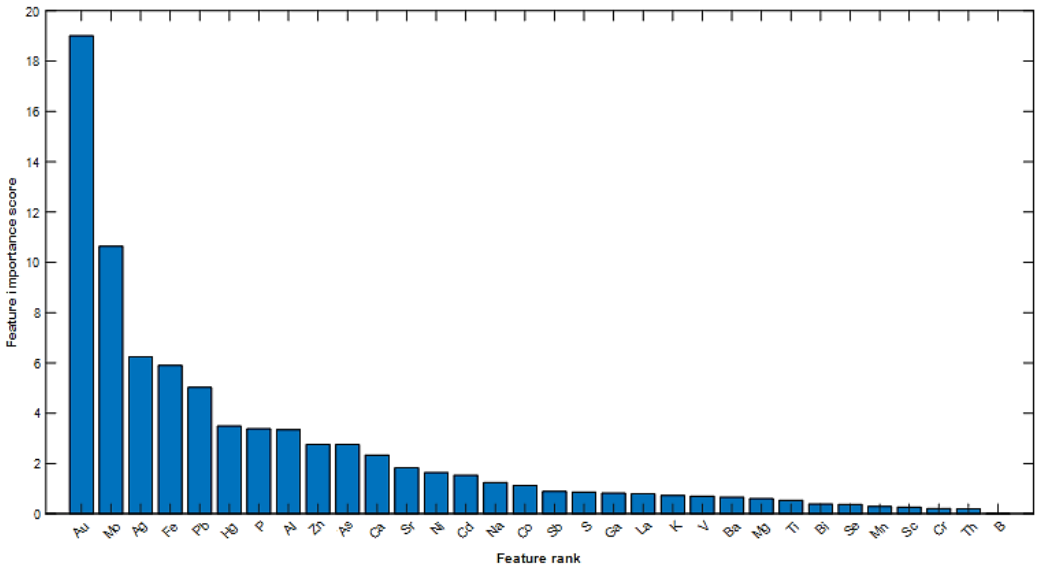


Fig. 3. Ranking of significant variables based on the F-test feature selection algorithm.

3 | Discussion and Results

The data were divided into training and testing datasets using an 80-20 split. In addition, 5-fold cross-validation ($k = 5$) was employed to validate the performance of the proposed Support Vector Regression (SVR) model. All simulations were carried out in MATLAB (R2022a) on a Windows 10 64-bit system equipped with an Intel® Core™ i7-11700K processor and 16 GB RAM. The results of the copper prediction model on the validation and test datasets using the proposed SVR method with all features and with the five significant features are presented in *Tables 2* and *3*, respectively. In these tables, Mean Squared Error (MSE), Root Mean Square Error (RMSE), $R^2$and Mean Absolute Error (MAE) represent MSE, RMSE, R-squared (R^2), and MAE, respectively, and are used to compare the results [8].

Table 2. Results of copper concentration prediction using the validation dataset.

Algorithm	MSE	RMSE	R^2	MAE	Training Time
Proposed SVR model with all features	0.0291	0.1707	0.06	0.0642	
Proposed SVR model with selected features	0.0077	0.0879	0.66	0.0441	1.34

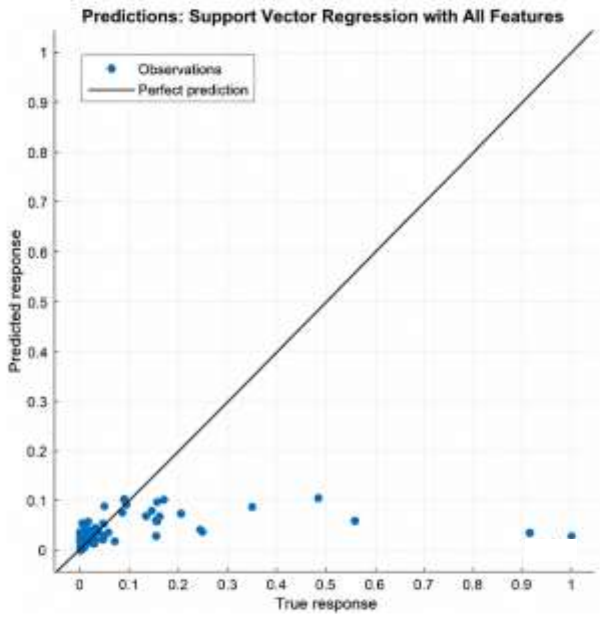
Table 3. Results of copper concentration prediction using the test dataset.

Algorithm	MSE	RMSE	R^2	MAE
Proposed SVR model with selected features	0.0109	0.1045	0.77	0.0545

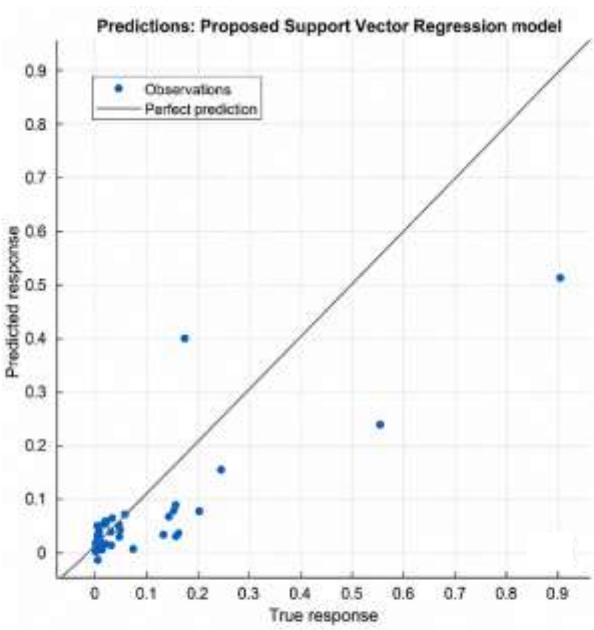
The measured and predicted copper concentration values for the validation and test datasets using the proposed SVR model with all features and with the five selected significant features are shown in *Fig. 4*. As can be observed in *Fig. 4.a*, the results of the SVR method using all features show a considerable discrepancy compared to the measured values, with a very low R^2 value of 0.06 and an RMSE error of 0.1701 [9].

On the other hand, the results of the SVR method using the selected features, shown in *Fig. 4.b*, demonstrate a substantial improvement, yielding a relatively acceptable R^2 value of 0.66 and an RMSE error of 0.088, which indicates the importance of feature selection using the F-test method. Furthermore, the results of the SVR method using the selected features on the test dataset are also shown in *Fig. 4.b*, with a favorable R^2 value of 0.77 and an RMSE of 0.0109, which is better than the validation dataset results obtained using the same model [9].

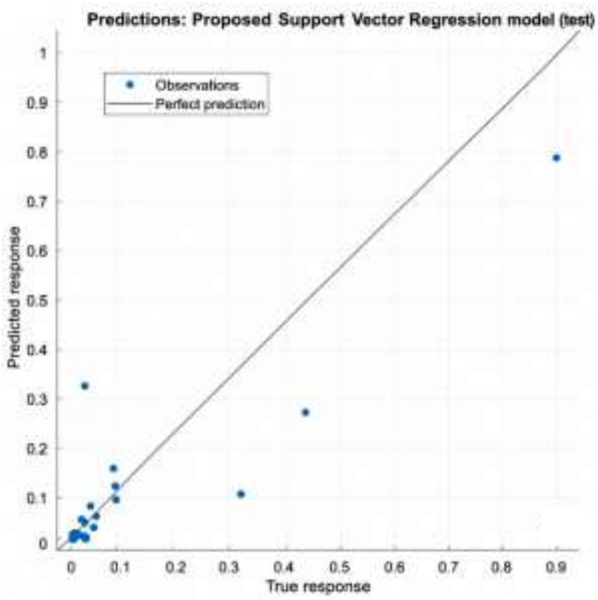
In addition, Fig. 5 illustrates the error of the SVR method using the selected features; it is visually interpretable and comparable, so it requires little further explanation.



a.



b.



c.

Fig. 4. Measured and predicted copper concentration values on the validation dataset using: a. the proposed SVR model implemented with all features, b. the SVR model implemented with the selected features, and c. measured and predicted copper concentration values on the test dataset using the SVR model implemented with the selected features.

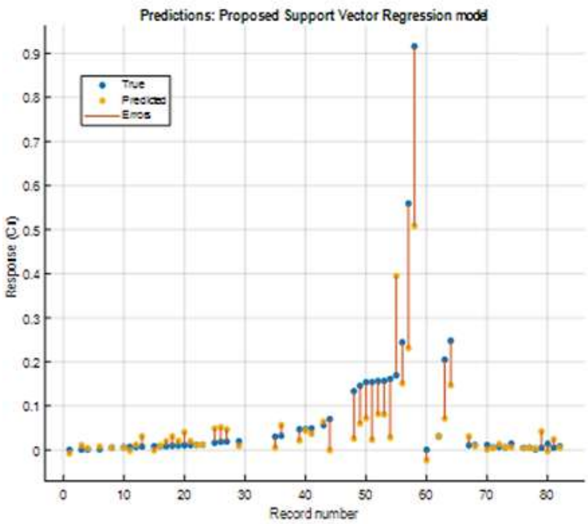
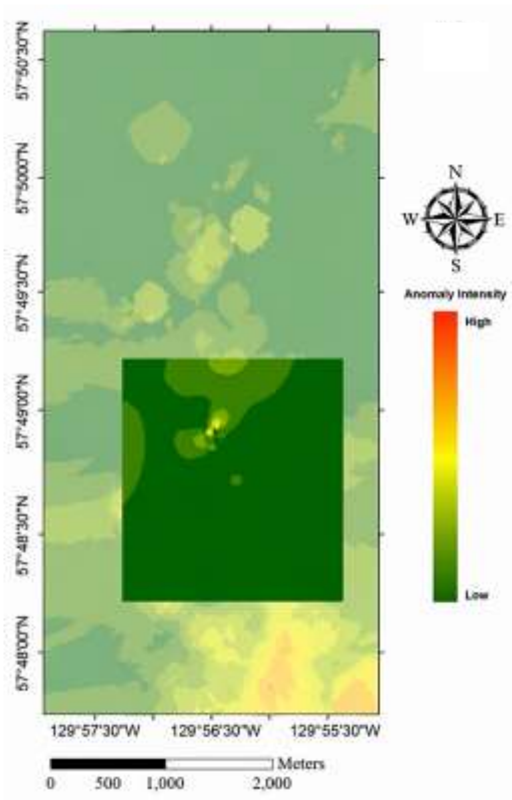
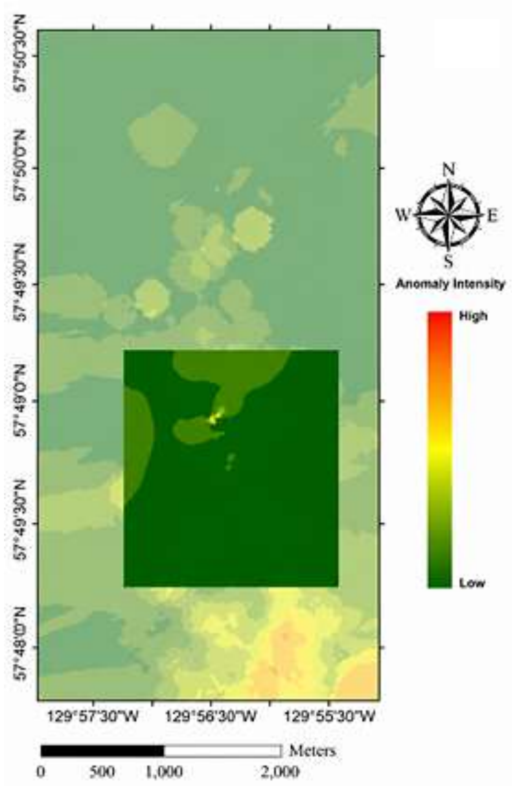


Fig. 5. Measured and predicted copper concentration values on the validation dataset using the SVR model with the selected features.

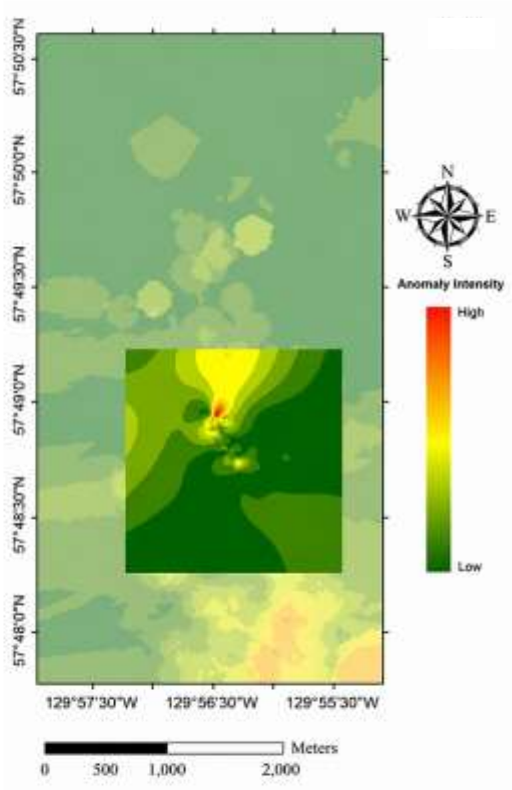
Fig. 6 .a shows the map of prospective copper zones based on the actual measured values from the rock samples, generated in ArcGIS. *Fig. 6.b* presents the prospective zone map produced using the proposed SVR model with all features, while Fig. 6 .c shows the prospective zone map generated using the proposed SVR model with the selected features Au, Mo, Ag, Fe, and Pb. The regression model results based on all features identify extensive areas as anomalous zones and differ considerably from the prospective zone map based on actual copper values (*Fig. 6.a*). In contrast, the map generated by the proposed SVR model using the selected features (*Fig. 6.c*) exhibits a high degree of correspondence, demonstrating the efficiency of the proposed regression method in predicting copper concentrations [5].

a.





b.



c.

Fig. 6. Maps of prospective copper zones based on: a. actual copper values obtained from rock samples, b. predicted copper values using the proposed SVR model with all features, and c. predicted copper values using the proposed SVR model with the selected features.

4 | Conclusion

This study aimed to identify prospective copper mineralization zones in a porphyry copper mineralization area in British Columbia, Canada, using rock sample data through geochemical investigation and ML methods. The geochemical analysis of the rock samples produced a dataset of 83 samples (rows) and 36 elements (variables/features, columns).

Subsequently, dimensionality reduction of the data matrix was performed using the F-test feature selection method, through which five significant variables were identified. Thereafter, copper concentration, as the target variable, was predicted based on the proposed SVR method using both all features (elements) and the selected features.

Finally, the results of each method were presented as prospective copper mineralization zone maps, demonstrating the efficiency of the SVR regression model using the selected features.

Authors' Contributions

All aspects of the research and manuscript preparation were carried out by the author. The author has read and approved the final version of the manuscript.

Funding

Not applicable.

Data Availability

All data supporting the reported findings in this research paper are provided within the manuscript.

Conflict of Interest

The author declares that they do not have any conflict of interest.

Consent for Publication

The author confirms consent for the publication of this work.

Ethics Approval and Consent to Participate

This article does not contain any studies with human participants performed by the author.

References

[1] Carranza, E. J. M. (2009). *Geochemical anomaly and mineral prospectivity mapping in GIS*. Elsevier. https://shop.elsevier.com/books/geochemical-anomaly-and-mineral-prospectivity-mapping-in-gis/carranza/978-0-444-51325-0

[2] Hosseini-Dinani, H., Mokhtari, A. R., Shahrestani, S., & De Vivo, B. (2019). Sampling density in regional exploration and environmental geochemical studies: A review. *Natural resources research*, *28*(3), 967–994. https://doi.org/10.1007/s11053-018-9431-2

[3] Zuo, R., Wang, J., & Xiong, Y. (2019). Deep learning and its application in geochemical mapping. *Earth-science reviews*, *192*, 1–14. https://doi.org/10.1016/j.earscirev.2019.02.023

[4] Zuo, R. (2017). Machine learning of mineralization-related geochemical anomalies: A review of potential methods. *Natural resources research*, *26*(4), 457–464. https://doi.org/10.1007/s11053-017-9345-4

[5] Xiong, Y., & Zuo, R. (2020). Recognizing multivariate geochemical anomalies for mineral exploration by combining deep learning and one-class support vector machine. *Computers & geosciences*, *140*, 104484. https://doi.org/10.1016/j.cageo.2020.104484

[6] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*, 273–297. https://doi.org/10.1007/BF00994018

[7] Bommert, A., Sun, X., Bischl, B., Rahnenführer, J., & Lang, M. (2020). Benchmark for filter methods for feature selection in high-dimensional classification data. *Computational statistics & data analysis*, *143*, 106839. https://doi.org/10.1016/j.csda.2019.106839

[8] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer. https://doi.org/10.1007/978-0-387-84858-7

[9] Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *Peerj computer science*, *7*, e623. http://dx.doi.org/10.7717/peerj-cs.623