Physics and Chemistry of the Earth 136 (2024) 103766

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

journal homepage: www.elsevier.com/locate/pce

Multi-method machine learning techniques in gold pathfinder elements prediction in central parts of Tanzania using stream sediment geochemical data

Samuel Nunoo [a], Mahamuda Abu [b,*], Emmanuel Ayitey [c], Benatus Norbert Mvile [d], John Desderius Kalimenze [e]

[a] Department of Earth Science, School of Physical and Mathematical Sciences, College of Basic and Applied Science, University of Ghana, P. O. Box LG 58, Legon, Ghana
[b] Department of Geological Engineering, School of Engineering, University for Development Studies, P. O. Box 1882, Nyankpala, Tamale, Ghana
[c] Department of Mathematics, Statistics and Actuarial Science, Takoradi Technical University, Box 256, Takoradi, Ghana
[d] Department of Physics, College of Natural and Mathematical Sciences, University of Dodoma, P. O. Box 259, Dodoma, Tanzania
[e] Geological Survey of Tanzania (GST), P. O. Box 903, Dodoma, Tanzania

ARTICLE INFO

Keywords:
Multi-technique
Machine learning
Exploration targeting
Gold
Tanzanian craton

ABSTRACT

Prediction models using machine learning techniques have proven to be a reliable technique in mineral exploration. A combination of these techniques is very robust and reliable in exploration targeting and much dependable as an approach in greenfield. In this study, multi-machine learning methods: random forest (RF), support vector machine (SVM), and artificial neural network (ANN) were employed to conduct a data-driven gold (Au) prospectivity modelling in the Central parts of the Tanzania Craton (TC). A total of 166 samples with Au concentrations from stream sediment samples were considered. Based on the modeling results, the RF model demonstrates superior prediction accuracy compared to the SVM (MSE of 0.89) and ANN models (MSE of 1.21), achieving an MSE of less than 0.82. In terms of overall predictive performance and efficiency, the RF model outperforms other ML models deployed in this research. Therefore, it is deemed the suitable model for gold (Au) prediction in the TC catchments. According to the geological interpretation derived from the model, anomalies in arsenic (As), nickel (Ni), and tungsten (W) now emerge as significant predictors in the quest for gold. This implies that the association of As–Ni–W are potential pathfinder elements in the exploration of gold in the central part of the TC.

1. Introduction

The advent of advanced technology and the rapid growth of humanity in population and industrialization means a continuous reliance on earth's natural resources, especially solid minerals. The exploration for earth materials such as rocks, minerals and elements dates over a century (e.g., Nunoo et al., 2023). Even in the brown fields, where commercial mineral resources and reserves exist, the search for additional and new minerals keeps growing (Cabello, 2021; Forson and Amponsah, 2023a; 2023b; Atanga et al., 2023). The challenge in exploration largely centers on where to focus in any given terrain, especially to locate any mineral or element of a promising content or anomalous concentration. The traditional approach often considers the content value of each sample point. This method has been fruitful in many ways but time consuming when dealing with large data sets, and consequently affects decision making that is crucial in mineral exploration. For instance, many have used various univariant to multivariant, and other artificial intelligence approaches such as machine learning techniques to handle large data set in the search and monitoring of geological resources such as rock, water and minerals (Nude et al., 2012; Forson; Amponsah 2023b; Nunoo et al., 2023; Abu et al., 2023, 2024a, 2024b). These methods do not only offer means to handle large data sets, but are effective in decision making, besides offering real-time robust data processing. As new data handling methods emerge, they indirectly become means of comparing with previous methods.

This research targets the prediction of gold mineralization and occurrence using well established pathfinder element in the exploration of gold. For example, elements such as arsenic (As), molybdenum (Mo),

* Corresponding author.
E-mail address: mahamudaabu@gmail.com (M. Abu).

https://doi.org/10.1016/j.pce.2024.103766
Received 29 July 2024; Received in revised form 6 October 2024; Accepted 15 October 2024
Available online 16 October 2024

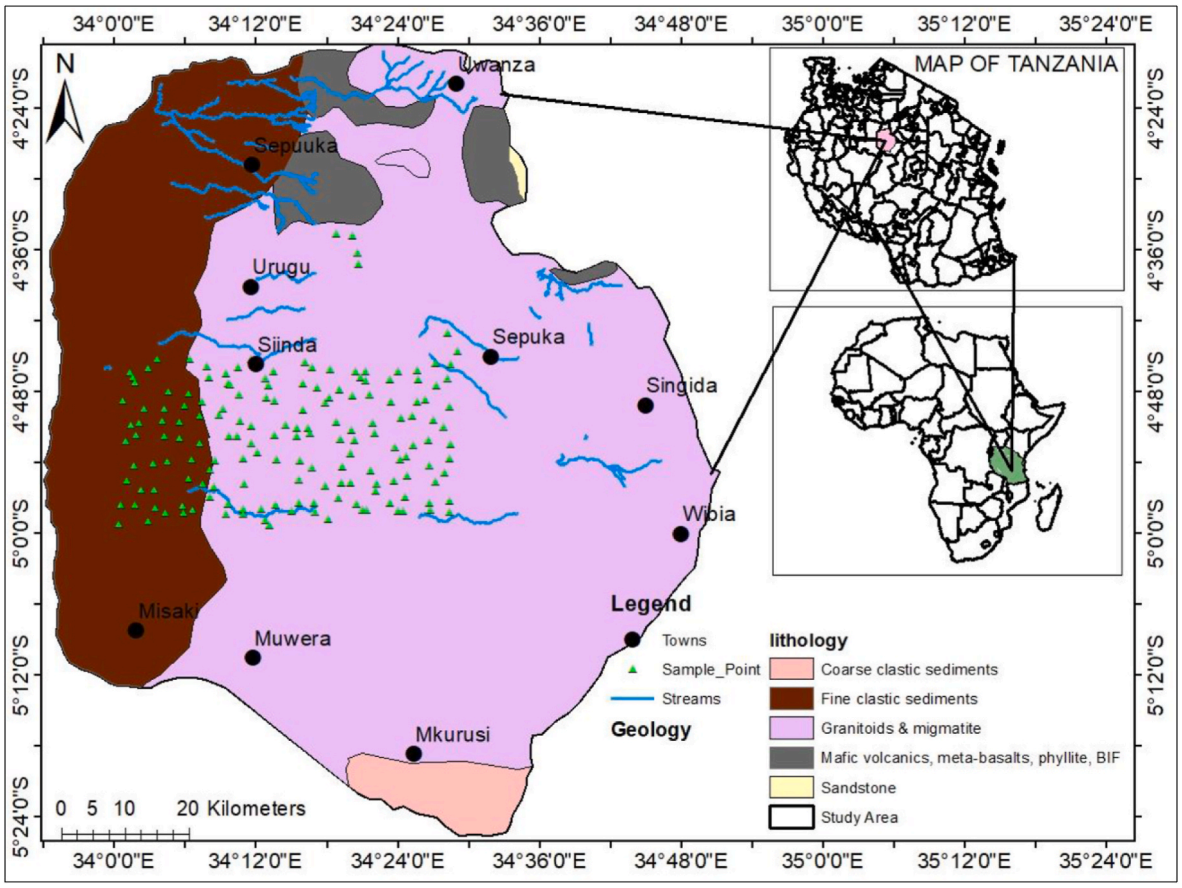

Fig. 1. Geological map of the study area with the sample locations (Nunoo et al., 2023).

copper (Cu), bismuth (Bi), tungsten (W), tin (Sn) and nickel (Ni) are potential pathfinders in the search for gold (e.g., Plouffe 2001; Bayari et al., 2019; Somarin et al., 2021). The degree of affinity of these elements to gold may not be entirely uniform, as their degree of correlation may depend on the nature of the geological terrain. For instance, elsewhere in the Birimian greenstone belt of Ghana, especially the southern part, the association of gold with As–S is a strongly known pathfinder (Allibone et al., 2002), but in the central and northern segment of the same country, others have reported additional and or other potential pathfinders to gold as Fe–Mg and Pb–Ag–As–Cu (e.g., Nude et al., 2012; Nzulu et al., 2023). Such variation in pathfinders to gold mineralization is suggestive of terrain uniqueness, and the need to perform rigorous or robust analysis to ascertain reliable pathfinder element to gold in any given terrain. Thus, this research seeks to use various machine learning algorithms to predict which of the pathfinder elements of As–Mo–Cu–Bi–W–Sn–Ni is the most reliable in the search of gold mineralization within the Tanzania Craton, especially in the central part of the craton. Previous work on mineralization in this part of the craton, used multivariant geostatistical methods and ordinary kriging interpolation methods to offer firsthand exploration targets on forty-three elements (Mvile et al., 2021; Nunoo et al., 2023; Abu et al., 2024a). The question is on how the approach adopted by Nunoo et al. (2023) compares with the emerging Artificial Intelligence (AI) methods, especially Machine Learning (ML) techniques. For instance, the work of Abu et al. (2024a) and Nunoo et al. (2023) from the Tanzania terrain show poor to no correlation of Au to other well-documented pathfinders (e.g., As, Cu). Could some AI methods within the ML algorithms offer any form of indicative pathfinders to gold? This remains to be explored.

For example, Omali (2021), Feizi et al. (2017) and Mathew and Ariffin (2018) have emphasized the applicability of ML technique in mineral prospecting (Forson et al., 2022).

The focus of the study was to predict the most reliable pathfinder element, when prospecting for gold occurrence within the study area, aspects which were less projected in previous study (e.g., Abu et al., 2024a; Nunoo et al., 2023), despite its importance in exploration targeting. Different machine learning algorithms of Random Forest (RF), Support Vector Machine (SVM) and Artificial Neural Network (ANN) were adopted in this research. The rational is to see the most efficient technique in the geological terrain being explored, and or to ascertain the complementarity of these emerging techniques. In addition, Variance Inflation Factor (VIF) and sensitivity analysis were incorporated to respectively estimate how regression coefficients is inflated due to multicollinearity in the independent variables and the change in the output variable relative to a change in the input variable. The study targets to enhance exploration success and decision making within the broad Tanzania Craton (TC) while providing options of data processing techniques that projects the true reflectivity of the data.

2. Geological setting

The area is within the broader TC and is located within the central part of Tanzania (Kabete et al., 2012). The TC has several greenstone belts including the cluster of greenstone belts characterized by felsic gneisses and granitic rocks together with some schists (Borg and Shackleton, 1997; Kabete et al., 2012). The Siinda-Sepuka-Misaki-Singida area (Fig. 1) is within the Iramba-Sekenke greenstone belt (Cook et al., 2015; Kwelwa et al., 2018), comprising granitoids and quartzo-felspathic gneisses, as well as stratified metavolcanic and metasedimentary rocks within the greenstone belt (Henckel et al., 2016). The older geological sequence is marked by felsic volcanic rock together with volcaniclastics, banded chert-oxide and iron formations associated with mafic volcanic rock unit (Henckel et al.,

Table 1
ICP-MS analytical data of gold (ppb) and selected pathfinder elements (ppm) obtained from stream sediment samples.

Sample ID	X	Y	Mo	Cu	Ni	As	Au	Bi	W	Sn
1012I4SSE	663013	9478520	1.20	**32.90**	**58.60**	2.00	0.81	0.20	1.70	1.00
1012I5SSE	664577	9475544	0.90	28.50	52.60	4.00	0.74	0.30	**5.20**	1.40
1012I6SSE	613400	9472355	0.60	8.50	9.10	1.00	0.67	0.30	1.40	1.80
1012I7SSE	613755	9471639	0.70	11.80	12.80	1.00	0.73	0.00	0.60	1.50
1012I8SSE	612197	9468059	0.90	7.00	6.70	1.00	0.85	0.00	0.50	1.60
1012I9SSE	613460	9464233	1.40	10.90	13.20	2.00	1.91	0.10	0.80	2.70
1013A1SSE	612686	9461683	0.70	4.30	5.10	0.00	1.78	0.00	0.70	0.60
1013A2SSE	613917	9457841	1.10	6.00	8.30	1.00	1.45	0.00	1.20	0.80
1013A3SSE	612811	9455576	0.80	2.30	2.70	0.00	2.27	0.00	0.60	0.50
1013A4SSE	611908	9451661	0.60	2.60	2.90	0.00	0.51	0.00	0.50	0.50
1013A5SSE	611615	9448670	1.60	5.30	8.10	0.00	0.40	0.00	1.10	0.90
1013A6SSE	616209	9472951	0.70	6.40	9.10	0.00	0.55	0.20	1.40	1.50
1013A7SSE	614220	9470791	0.60	5.80	6.30	0.00	0.55	0.10	1.00	1.00
1013A8SSE	615541	9466753	0.90	7.10	10.50	0.00	1.71	0.20	1.50	1.50
1013A9SSE	614810	9464478	0.60	7.00	10.10	2.00	0.77	0.10	1.10	1.20
1013B1SSE	614274	9462628	1.00	5.70	6.80	0.00	0.51	0.00	0.80	0.80
1013B2SSE	616960	9458161	0.70	8.30	14.40	0.00	0.45	0.10	1.40	1.40
1013B3SSE	615080	9454092	0.30	3.40	4.40	0.00	3.69	0.00	1.00	0.80
1013B4SSE	614138	9451798	0.60	4.40	8.10	0.00	0.82	0.20	1.40	1.50
1013B5SSE	616168	9449244	1.00	6.00	10.70	0.00	1.02	0.10	1.40	1.60
1013B6SSE	617615	9474437	0.50	4.60	7.40	1.00	0.58	0.10	0.80	1.00
1013B7SSE	619316	9469458	0.90	7.70	11.80	1.00	0.47	0.20	1.40	1.50
1013B8SSE	645742	9493986	0.50	5.50	10.40	0.00	0.76	0.10	1.30	1.10
1013B9SSE	618771	9466775	1.30	9.70	16.00	1.00	0.73	0.20	1.30	1.60
1013C1SSE	618382	9464872	0.80	12.60	15.20	0.00	0.34	0.10	1.20	1.80
1013C2SSE	618643	9462443	1.20	9.40	18.70	1.00	1.55	0.20	1.30	2.70
1013C3SSE	619136	9458436	0.90	7.40	7.30	2.00	0.74	0.10	0.80	1.30
1013C4SSE	617053	9454033	0.60	2.40	2.90	0.00	0.75	0.00	0.60	0.40
1013C5SSE	617207	9451178	0.90	9.90	12.60	0.00	1.04	0.20	1.20	2.10
1013C6SSE	618805	9450493	0.70	10.70	15.10	1.00	0.54	0.20	1.50	1.90
1013C7SSE	622707	9474384	1.50	13.20	22.20	0.00	0.56	0.20	2.40	2.50
1013C8SSE	622507	9469106	0.60	6.30	8.80	1.00	0.66	0.20	1.10	1.10
1013C9SSE	621819	9467113	0.50	7.30	9.40	3.00	0.68	0.10	1.10	1.30
1013D1SSE	648294	9493670	**4.10**	13.70	19.90	2.00	0.67	0.20	2.70	3.30
1013D2SSE	649043	9490950	1.70	9.30	18.70	3.00	0.64	0.20	1.20	2.60
1013D3SSE	621127	9464579	0.90	4.80	8.50	1.00	0.90	0.00	0.80	1.20
1013D4SSE	649120	9489238	0.60	3.80	5.60	2.00	0.71	0.00	0.60	1.20
1013D5SSE	621126	9462182	1.10	2.40	3.70	2.00	0.74	0.00	0.60	0.40
1013D6SSE	622976	9458806	0.90	3.40	4.90	0.00	0.68	0.00	0.70	0.80
1013D7SSE	622525	9455778	0.90	15.40	23.10	3.00	0.51	0.30	1.90	2.60
1013D8SSE	621624	9451499	1.00	8.00	12.90	2.00	0.85	0.20	1.20	1.80
1013D9SSE	621720	9450513	0.70	5.80	7.90	1.00	0.46	0.20	1.00	1.30
1013E1SSE	625853	9457188	0.70	4.90	7.50	2.00	0.54	0.00	0.60	1.10
1013E2SSE	624745	9455084	0.70	5.70	10.20	2.00	1.07	0.10	0.60	1.70
1013E3SSE	625885	9452922	1.40	12.70	26.80	3.00	0.61	0.20	1.50	3.40
1013E4SSE	623079	9450949	0.50	8.50	10.90	1.00	0.42	0.10	0.70	1.60
1013E5SSE	625350	9473363	1.40	11.00	16.00	2.00	0.59	0.30	1.40	2.40
1013E6SSE	625734	9471637	0.20	3.80	5.60	0.00	0.69	0.10	0.40	1.00
1013E7SSE	624592	9467775	1.10	13.60	16.50	1.00	0.69	0.20	1.40	2.20
1013E8SSE	624260	9465660	1.60	13.90	23.00	1.00	0.59	0.30	2.00	2.90
1013E9SSE	624220	9461400	1.20	13.10	18.50	2.00	0.40	0.30	1.70	2.40
1013F1SSE	628435	9472173	0.80	5.30	7.90	2.00	0.42	0.10	0.60	1.20
1013F2SSE	628735	9470632	1.00	7.40	13.60	2.00	0.40	0.10	0.90	2.70
Sample ID	X	Y	Mo	Cu	Ni	As	Au	Bi	W	Sn
1013F3SSE	627929	9466651	0.90	6.70	10.80	2.00	0.43	0.20	1.00	2.00
1013F4SSE	627340	9465755	0.60	5.40	7.20	0.00	1.19	0.20	0.70	1.60
1013F5SSE	628813	9462409	0.50	2.20	3.10	1.00	**4.34**	0.00	0.40	0.60
1013F6SSE	626672	9458591	0.70	6.10	10.70	0.00	1.55	0.20	0.70	1.80
1013F7SSE	626490	9454560	1.40	12.30	20.10	0.00	1.12	0.20	1.20	2.20
1013F8SSE	628618	9451894	0.80	9.20	12.60	0.00	1.51	0.20	0.90	1.70
1013F9SSE	628433	9450305	1.40	18.20	29.50	0.00	0.51	0.20	2.00	3.00
1013G1SSE	630256	9473225	0.30	2.80	5.10	0.00	0.42	0.10	0.50	1.60
1013G2SSE	629065	9470492	0.80	3.40	4.90	0.00	0.26	0.00	0.60	1.40
1013G3SSE	630487	9468833	1.10	9.70	9.00	0.00	0.30	0.10	0.80	2.30
1013G4SSE	630657	9464730	0.30	2.50	2.60	1.00	0.49	0.00	0.30	0.90
1013G5SSE	630157	9462399	0.90	2.00	3.10	0.00	0.53	0.10	0.10	1.20
1013G6SSE	630992	9458919	3.40	15.00	26.20	2.00	0.54	0.30	1.70	3.50
1013G7SSE	631192	9454981	1.70	8.30	14.80	2.00	0.51	0.10	0.90	1.80
1013G8SSE	631022	9451099	0.70	4.60	9.70	0.00	0.39	0.00	0.80	1.10
1013G9SSE	630988	9450234	1.90	11.80	22.70	2.00	0.63	0.20	2.70	2.70
1013H1SSE	632281	9474374	0.50	3.00	3.00	1.00	0.47	0.20	1.00	2.20
1013H2SSE	634626	9471346	0.80	2.70	3.60	0.00	0.38	0.20	0.70	1.90
1013H3SSE	634772	9468494	0.70	4.90	9.50	0.00	0.34	0.00	0.50	1.30

(*continued on next page*)

Table 1 (*continued*)

Sample ID	X	Y	Mo	Cu	Ni	As	Au	Bi	W	Sn
1013H4SSE	632378	9463994	0.30	1.50	1.90	0.00	1.01	0.00	0.50	1.00
1013H5SSE	632305	9461402	0.60	2.50	2.90	0.00	0.47	0.20	0.70	1.50
1013H6SSE	633270	9459922	0.50	5.00	7.70	0.00	0.35	0.20	0.40	1.30
1013H8SSE	633960	9451129	1.80	10.80	19.10	0.00	0.37	0.20	1.50	3.10
1013H9SSE	634532	9449271	1.10	9.40	12.30	1.00	0.39	0.20	1.20	2.10
1013I1SSE	635761	9472588	1.00	5.40	8.30	0.00	0.56	0.30	0.90	2.60
1013I2SSE	635064	9470380	0.60	4.80	6.90	0.00	1.72	0.40	1.00	2.80
1013I3SSE	635977	9467969	0.60	3.30	4.90	0.00	0.33	0.00	0.70	2.00
1013I4SSE	637734	9464135	0.80	3.30	4.00	0.00	0.50	0.10	0.70	1.90
1013I5SSE	635974	9462652	0.50	4.40	5.20	0.00	0.56	0.00	0.40	0.80
1013I6SSE	636293	9458813	0.30	3.00	4.00	0.00	0.33	0.10	0.50	2.00
1013I8SSE	636178	9451651	0.80	8.90	18.30	0.00	0.31	0.10	0.90	2.30
1013I9SSE	635179	9448521	0.70	3.20	6.00	0.00	0.34	0.20	1.00	1.10
1014A1SSE	640669	9473892	0.80	4.40	5.10	0.00	0.32	0.30	3.20	**4.80**
1014A2SSE	640341	9470596	0.50	4.10	12.60	0.00	0.32	0.20	3.20	3.40
1014A3SSE	640404	9466495	0.40	3.60	3.60	0.00	0.28	**0.50**	0.70	1.50
1014A4SSE	639482	9463689	0.70	4.00	5.60	2.00	0.40	0.00	0.50	0.70
1014A5SSE	639402	9462450	0.40	2.40	6.20	0.00	0.38	0.00	0.30	0.70
1014A6SSE	639915	9458686	0.40	4.10	7.10	0.00	0.33	0.00	0.50	2.00
1014A7SSE	639818	9455557	0.90	4.90	8.40	1.00	0.69	0.00	0.60	1.40
1014A8SSE	639754	9452033	0.60	5.60	7.90	0.00	0.85	0.10	0.90	1.10
1014A9SSE	638659	9450728	0.40	7.20	12.20	0.00	1.78	0.30	1.30	1.90
1014B1SSE	641890	9473162	0.90	5.10	6.80	1.00	1.15	0.00	1.40	4.70
1014B2SSE	642691	9469990	0.60	1.90	3.60	0.00	1.03	0.00	1.40	2.20
1014B3SSE	643902	9467650	1.20	1.90	3.20	0.00	0.74	0.10	0.90	2.80
1014B4SSE	641139	9463641	0.70	2.30	5.70	2.00	0.69	0.20	0.70	1.60
1014B5SSE	641555	9462880	0.80	3.30	5.10	2.00	0.61	0.20	0.70	2.40
1014B6SSE	641845	9457576	0.60	7.10	9.80	0.00	0.76	0.40	0.60	2.70
1014B7SSE	643875	9454697	0.30	4.00	7.80	0.00	4.26	0.00	0.50	1.20
1014B8SSE	642089	9451970	1.70	11.40	18.60	1.00	0.56	0.20	2.10	2.80
1014B9SSE	642315	9450629	1.20	3.00	4.40	0.00	1.03	0.00	1.10	0.60
1014C1SSE	644685	9472770	0.50	5.50	6.40	0.00	1.53	0.00	0.50	1.10
1014C2SSE	646049	9469614	0.50	3.80	8.50	0.00	0.72	0.20	0.50	2.60
1014C3SSE	644491	9468409	1.00	2.90	5.10	0.00	0.75	0.00	0.60	1.30
1014C4SSE	645564	9464626	0.70	7.10	8.20	0.00	0.60	0.00	0.40	1.10
Sample ID	X	Y	Mo	Cu	Ni	As	Au	Bi	W	Sn
1014C5SSE	646189	9461997	0.90	3.10	3.40	0.00	0.58	0.00	0.60	0.50
1014C6SSE	646627	9457599	1.00	9.40	18.00	0.00	0.68	0.10	0.70	2.00
1014C7SSE	646357	9455182	0.80	6.40	14.60	0.00	1.11	0.10	0.70	1.60
1014C8SSE	645917	9452741	0.70	4.80	6.50	0.00	0.64	0.30	1.60	1.10
1014C9SSE	644273	9449430	1.90	16.60	28.50	0.00	1.32	0.40	2.40	3.40
1014D1SSE	648591	9472469	0.70	6.70	5.40	2.00	0.68	0.10	0.50	1.60
1014D2SSE	649524	9471560	1.10	6.40	5.50	0.00	0.74	0.10	1.30	4.30
1014D3SSE	648163	9468776	0.70	5.00	7.50	0.00	0.78	0.10	0.50	2.80
1014D4SSE	647990	9463374	0.50	5.80	7.00	0.00	0.80	0.00	0.40	1.00
1014D5SSE	648468	9461697	0.70	4.40	6.10	0.00	0.93	0.00	0.50	0.70
1014D6SSE	647876	9459743	0.50	5.10	8.60	0.00	0.95	0.10	0.50	1.20
1014D7SSE	647812	9455102	1.60	8.30	18.50	0.00	0.78	0.20	1.20	3.00
1014D8SSE	648820	9451871	3.20	17.80	30.90	3.00	0.80	0.30	2.00	4.10
1014D9SSE	649657	9450848	0.80	5.30	10.80	2.00	0.70	0.20	1.00	1.60
1014E1SSE	650248	9472513	1.20	5.70	7.40	0.00	0.69	0.00	0.70	2.00
1014E2SSE	650289	9471119	0.80	1.90	3.30	0.00	0.67	0.00	0.60	1.00
1014E3SSE	650799	9468879	0.60	5.90	6.70	0.00	1.05	0.10	0.70	3.60
1014E4SSE	651304	9464517	0.40	3.90	5.20	0.00	0.94	0.10	0.60	2.70
1014E5SSE	651660	9462794	0.30	6.30	6.70	0.00	0.93	0.00	0.40	1.80
1014E6SSE	650979	9457079	0.50	5.40	8.10	1.00	0.83	0.00	0.60	1.80
1014E7SSE	651923	9455022	0.60	5.90	12.50	0.00	0.92	0.00	0.60	1.70
1014E8SSE	651750	9453918	0.90	6.10	11.20	2.00	0.82	0.00	0.70	1.20
1014E9SSE	650411	9450900	0.40	1.90	3.90	0.00	0.73	0.20	1.00	1.20
1014F1SSE	655379	9472437	0.60	11.60	11.80	0.00	0.64	0.00	0.60	1.40
1014F2SSE	653965	9471199	0.60	3.80	5.70	0.00	0.83	0.00	0.50	1.00
1014F3SSE	654197	9467786	0.40	4.90	5.50	0.00	0.72	0.00	0.40	2.00
1014F4SSE	655718	9465161	0.90	13.10	18.60	0.00	1.03	0.10	0.90	3.50
1014F5SSE	654788	9460713	0.80	6.50	11.30	1.00	0.80	0.10	0.70	3.10
1014F6SSE	655520	9458741	0.50	5.60	7.40	0.00	0.78	0.10	0.70	1.70
1014F7SSE	654711	9455421	0.70	4.70	9.30	0.00	0.61	0.00	0.50	1.50
1014F8SSE	654067	9452012	0.70	7.80	9.70	0.00	0.74	0.20	0.50	1.70
1014F9SSE	655249	9450683	0.60	6.30	10.10	0.00	0.80	0.10	0.50	1.40
1014G1SSE	658282	9472698	1.90	17.40	25.30	**4.00**	0.86	0.20	1.30	4.60
1014G2SSE	658407	9471259	2.00	11.10	10.10	2.00	1.03	0.00	0.90	3.80
1014G3SSE	658319	9467719	0.50	3.50	3.70	0.00	0.73	0.00	0.40	1.70
1014G4SSE	657530	9465564	1.30	3.00	4.30	2.00	0.38	0.00	0.70	0.60
1014G5SSE	657625	9461007	0.60	5.70	6.40	1.00	1.66	0.10	0.70	2.10
1014G6SSE	656622	9459244	0.50	7.40	8.00	0.00	0.89	0.10	0.70	3.30

(*continued on next page*)

Table 1 (*continued*)

Sample ID	X	Y	Mo	Cu	Ni	As	Au	Bi	W	Sn
1014G7SSE	656741	9456193	0.50	8.30	13.00	2.00	0.66	0.00	0.50	1.40
1014G8SSE	656402	9452956	0.90	10.90	23.30	2.00	0.95	0.10	0.80	2.10
1014G9SSE	656038	9450629	0.50	7.10	10.50	2.00	1.93	0.00	0.80	1.60
1014H1SSE	661127	9473757	0.60	5.90	5.70	0.00	0.87	0.00	0.40	1.40
1014H2SSE	661383	9470208	1.00	4.30	4.80	0.00	0.90	0.00	0.60	1.70
1014H3SSE	659541	9468012	0.50	6.60	7.60	2.00	0.83	0.00	0.50	1.80
1014H4SSE	660781	9464826	0.50	5.00	6.30	2.00	0.84	0.00	0.50	2.00
1014H5SSE	660013	9461708	1.40	8.00	12.50	3.00	0.66	0.20	1.20	3.20
1014H6SSE	660243	9457778	0.60	7.70	12.80	2.00	0.36	0.00	0.70	2.60
1014H7SSE	661099	9455294	0.70	8.10	9.80	1.00	1.41	0.00	0.50	1.20
1014H8SSE	660099	9451906	1.30	12.00	22.60	2.00	0.87	0.00	0.70	1.50
1014H9SSE	660265	9450806	1.10	10.40	14.90	2.00	1.00	0.00	0.90	1.70
1014I1SSE	663602	9473515	1.50	11.10	12.70	3.00	1.08	0.00	1.00	3.10
1014I2SSE	663116	9471383	0.90	3.00	2.70	0.00	0.67	0.00	0.70	2.10
1014I3SSE	663444	9467001	1.00	8.90	12.00	3.00	1.11	0.00	0.70	3.30
1014I4SSE	662418	9463832	0.40	4.30	6.40	1.00	0.62	0.20	0.60	4.20
1014I5SSE	663584	9461033	1.30	12.60	19.30	2.00	0.88	0.20	1.30	3.80
1014I6SSE	663385	9458924	0.80	5.70	9.70	2.00	0.81	0.10	0.70	1.70
1014I7SSE	662534	9456633	0.70	7.20	11.30	2.00	0.63	0.00	0.60	1.60
1014I8SSE	663395	9452284	0.90	16.60	15.80	1.00	1.87	0.20	0.70	2.10
1014I9SSE	663407	9450573	0.70	7.80	14.10	1.00	1.74	0.10	0.70	1.70

2016). The intrusive rocks are dominantly K-feldspar-rich granitoids, which occur as batholiths throughout the belt (Chamberlain and Tosdal 2007). The mafic rock suite is largely foliated with gneisses and augen gneisses (Henckel et al., 2016).

3. Method and materials

3.1. Field sampling and preparation

One-hundred and sixty-six samples (Table 1) were collected via grab sampling method from streams within the study area (Fig. 1). Sampling was done using a small clean spade. Before sampling, organic matter at depositional sites of sediments was removed (to prevent sample contamination). Then a sample was collected from the B-horizon (10–20 cm deep), and the geographic position system (GPS) location was marked. The samples were collected into calico bags and then packed into plastic bags and kept in wooden boxes before being transported to the GST laboratory. In addition, to ensure that the sample condition was not obliterated during transportation, a sample temperature in the range of 27–31 °C was adhered to. Preliminary sample preparation of drying and sieving to < 2 mm fraction, was carried out in the field and later transported to the Tanzania Geological Survey where they were further sieved to 75 μm at the sample preparation section of the survey department.

3.2. Laboratory analysis

Five grams of the sieved samples (of 75 micron-sized grains) from each prepared sample in section 3.1 was clean-paper sealed and submitted to the Canadian Acme Laboratory for both major and trace elements analysis using the Inductively Couple Plasma Mass Spectrometry (ICP-MS) method. The platinum, palladium, and gold analyses were conducted at the Geological Laboratory of Henan, China. A thio-urea pre-concentrated of 10 g prepared from each sample after aqua regia decomposing, produced a residue that was dissolved again in aqua regia, and then the leached Au, Pd, and Pt were analyzed using ICP-MS. The analysis was conducted at a detection limit of 0.01 % for the oxides and 2 ppm for the trace elements. To ensure the quality of the results and the analytical procedure, duplicate and blank samples totaling 25, were inserted in the samples and submitted to the laboratories for analysis. The results of the 166 samples used in the study and the 25 samples for quality check purposes were repetitive with very minor differences. The analytical procedure followed the analytical protocols of Mvile et al. (2021) and Kalimenze et al. (2023). The applicability of ICP-MS method covers wider geoscience-related investigation, such as measuring the extent of metal content in the soil and water samples (e.g., Yuksel and Arica, 2018).

3.3. Variance Inflation Factor

The Variance Inflation Factor (VIF) is a measure used to detect multicollinearity in regression analysis. It assesses how much the variance of the estimated regression coefficients is inflated due to multicollinearity in the independent variables. The mathematical equation for VIF for the i^{th} independent variable is:

$$\boldsymbol{VIF_i} = \frac{\mathbf{1}}{\mathbf{1} - \boldsymbol{R_i}^2} \tag{1}$$

Where:

VIF_i is the Variance Inflation Factor for the i^{th} independent variable.

R_i^2 is the R^2 value obtained by regressing the i^{th} independent variable on all other independent variables. The VIF quantifies how much the variance of the estimated regression coefficient is inflated due to multicollinearity. Generally, a VIF value greater than 10 is considered indicative of significant multicollinearity.

3.4. Sensitivity of features variable

The coefficient of sensitivity, often denoted as S, represents the change in the output variable relative to a change in the input variable. It's typically expressed as the ratio of the percentage change in the output to the percentage change in the input.

Mathematically, the coefficient of sensitivity (S) can be expressed as

$$\boldsymbol{s} = \frac{\%\ \boldsymbol{change\ in\ output}}{\%\ \boldsymbol{change\ in\ input}} \tag{2}$$

If y represents the output variable and x represents the input variable, and we denote the initial values of y and x as y_0 and x_0 respectively, and the new values after a change as y_1 and x_1, then:

$$\%\ \boldsymbol{change\ in\ output} = \frac{\boldsymbol{y_1 - y_0}}{\boldsymbol{y_0}} \times \mathbf{100\%} \tag{3}$$

$$\%\ \boldsymbol{change\ in\ input} = \frac{\boldsymbol{x_1 - x_0}}{\boldsymbol{x_0}} \times \mathbf{100\%} \tag{4}$$

Therefore, the coefficient of sensitivity can also be written as:

$$S = \frac{y_1 - y_0}{y_0} \times \frac{x_0}{x_1 - x_0} \tag{5}$$

This equation gives the coefficient of sensitivity, representing how much the output changes relative to a change in the input.

3.5. Machine learning (ML) algorithm

3.5.1. Random forest (RF)

One popular algorithm for group learning is Random Forest, which combines multiple decision trees to generate repeated predictions of the phenomena represented in the training dataset (Breiman et al., 2001). Decision trees form the foundational classifiers within the "forest," each trained on various subsets of the training data. These subsets are created using a sampling technique called bagging, where training subsets are generated by randomly selecting data points from the original dataset with replacement (Galiano and Rodriguez 2015). In Random Forests, trees grow by repeatedly splitting root nodes into binary leaf nodes. This process continues at each internal node until a predetermined stopping condition is reached (Carranza et al. 2015). Unlike regular decision trees, RF randomly selects subsets of input predictor variables at each node in the tree ensemble to serve as discriminative conditions.

Next, the RF algorithm evaluates each split to maximize the purity of the resulting trees. Purity measures how accurately a randomly selected sample from the input dataset would be labeled according to the distribution of labels in subsets. The Gini index (IG), commonly used in this context, quantifies the information purity of leaf nodes compared to their root nodes, expressed as follows (Breiman et al., 1984):

$$I_G(f) = \sum_{i=1}^{n} f_i(1 - f_i) \tag{6}$$

where f_i is the probability of class i at node n. This is defined as: $I_G(f) = \sum_{i=1}^{n} f_i(1 - f_i)$.

$$f_i = \frac{m_i}{m} \tag{7}$$

In RF, m_i represents the number of samples in class j, and m denotes the total number of samples in a node. The final prediction of RF is determined by majority voting across all decision trees. Breiman et al. (2001) suggest that random sampling and variable selection reduce the correlation between individual trees, enhancing forest diversity and boosting algorithm robustness while mitigating overfitting risks.

3.5.2. Support vector machine (SVM)

Support Vector Machines (SVMs) are a powerful supervised learning algorithm used for classification and regression tasks. They are based on the concept of finding an optimal hyperplane that maximizes the margin between different classes in the training data (Asadi et al., 2001). SVMs are effective in high-dimensional spaces and are particularly useful when the number of dimensions exceeds the number of samples. They can handle non-linear relationships using kernel functions, which transform input data into higher-dimensional spaces where a linear separation can be achieved. SVMs are known for their ability to generalize well and their robustness against overfitting, making them widely used in various fields such as image classification, bioinformatics, and text categorization (Vapnik et al., 2000).

For regression tasks, SVMs aim to find a function that predicts continuous values (Zhang et al., 2018). The basic formulation can be summarized as follows: Given a training dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, where $\mathbf{x}_i \in \mathbb{R}^p$ are the input features and $y_i \in \mathbb{R}^p$ are the corresponding outputs, the SVM regression seeks to find a function f(x) that minimizes the empirical risk while penalizing deviations from the predicted values (Huang et al., 2002):

$$\min_{f,\epsilon_i \geq 0} \left(\frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \epsilon_i \right)$$

subject to.

$$y_i - f(\mathbf{x}_i) \leq \epsilon_i$$

$$f(\mathbf{x}_i) - y_i \leq \epsilon_i$$

Here:

$$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$$

$f(\mathbf{x}) = \langle \mathbf{w}, \mathbf{x} \rangle + b$ is the regression function where $\mathbf{w}$ are the weights, b is the bias term, and ϵ_i are slack variables that allow for deviations from the exact prediction. C is a regularization parameter controlling the trade-off between minimizing the norm of $\mathbf{w}$ and minimizing the deviations ϵ_i.The inequalities ensure that the predictions $f(x_i)$ lie within a margin of y_i, modulated by ϵ_i.In practice, SVM regression often uses a kernel function K (x_i,y_j) to implicitly map the input space into a higher-dimensional feature space, allowing for non-linear relationships to be captured (Zuo et al. 2011).

3.5.3. Artificial neural network (ANN)

Artificial Neural Networks (ANNs) are computational models inspired by biological neural networks, designed to learn complex mappings from input data to continuous output values (Zaremotlagh et al., 2017). They consist of interconnected layers of neurons. In an ANN, the output $\widehat{y}$ is computed by sequentially passing inputs through layers (Brown et al., 2000):

$$\widehat{y} = f^{(L)}\left(\mathbf{W}^{(L)} f^{(L-1)}\left(\ldots f^{(1)}\left(\mathbf{W}^{(1)}\mathbf{x} + \mathbf{b}^{(1)}\right) + \mathbf{b}^{(L-1)}\right) + \mathbf{b}^{(L)}\right)$$

Sigmoid, tanh, or ReLU introduce non-linearities crucial for learning complex relationships in data. ANNs are trained via backpropagation, adjusting weights and biases to minimize a loss function (e.g., Mean Squared Error) between predicted $\widehat{y}$ and actual outputs y (Celik et al., 2017). Number of layers L, neurons per layer, activation functions, learning rate, and batch size impact model performance and training efficiency (Rodriguez-Galiano et al., 2015).

3.6. Model training

Training procedures involved the preparation of input data, followed by the development of machine learning models. A key focus was placed on identifying crucial parameters utilized by these models. In data-driven modeling, selecting appropriate parameter configurations can prove challenging as optimal settings vary based on the dataset and application context. No singular empirical guideline exists for determining these parameters; instead, a methodical trial-and-error approach is essential. The prediction performance of trained machine learning models was assessed through cross-validation, considering various parameter combinations. Evaluation metrics included the mean squared error (MSE).

3.7. Model assessment

The prediction performance of the trained models was assessed using the mean squared error (MSE), calculated as follows:

$$MSE = \frac{1}{N} \sum (\widehat{y}_i - y_i)^2 \tag{16}$$

where N represents the total number of samples in the test dataset, $\widehat{y}_i$ denotes the expected outcome for each sample, and y_i represents the true value of each sample. The optimal model configuration refers to the set of parameters that, when trained, produces the best model with the lowest mean squared error.

Table 2
Parameters used for training machine learning models.

Model	Parameter	Description	References
RF	Number of trees	Number of trees in the random forest	10–500
	Number of features	Number of features used for developing each tree	1–8
SVM	Gamma	A width parameter of RBF that determines the influencing range of each support vector	0.1–1
ANN	Number of neurons	Number of neurons in the hidden layer	2–15
	Learning rate	Change rate of weight in the training process	0.1–0.5

Table 3
Summary statistics of Au and selected pathfinders.

Variables	Mean	Max.	Min.	Std.	Skewness
Mo	0.881	4.100	0.200	0.524	***8.245***
Cu	7.080	32.9 00	1.5000	4.503	***5.312***
Ni	10.740	58.600	1.900	7.869	0.356
As	0.813	4.000	0.000	1.018	0.564
Bi	0.110	0.500	0.000	0.109	***5.523***
W	0.959	5.200	0.100	0.626	1.123
Sn	1.877	4.800	0.400	0.927	***7.234***
Au	0.839	4.340	0.258	0.584	0.783

4. Results

Generally, the main result has been presented in Table 1. Such result forms the basis for from which the descriptive statistic summarized in Table 3 evolved, besides the interpolation maps further produced in this research. In Table 1, the elements of investigation (i.e. Au, Ni, Cu, As, Bi, W, Sn and Mo) the highest values and the sample with such content have been highlighted. For example, highest value of Cu (32.90 ppm) and Ni (58.60 ppm) were both recorded in same sample (ID: 1012I4SSE). The highest value of Au (4.34 ppb) occurs in sample 1013F5SSE, which recorded zero value for Bi. Other elements of Mo, As, Bi, W and Sn recorded highest values of 4.10 ppm, 4.00 ppm, 0.50 ppm, 5.20 ppm and 4.80 ppm in the respective samples of 1013D1SSE, 1014G1SSE, 1014A3SSE, 1012I5SSE and 1014A1SSE. A further keen observation indicates that sample IDs of 1012I are potential ideal for the occurrence of Cu, Ni and W, whiles IDs of 1014A projects presence of Bi and Sn, other aspects of the data in terms of sensitivity measure and the degree of inter-correlation have been further explained. These elements distribution and or enrichment are mostly geologically induced via processes of weathering, erosion and re-deposition, rock-water interaction and other mechanisms yet to be understood (e.g., McLennan et al., 1993; Fedo et al., 1996).

Table 4
Correlation analysis between dependent and independent variables.

Dependent Variable	Independent Variables	Correlation Coefficient
Au	Mo	0.384
	Cu	0.251
	Ni	**0.813**
	As	**0.685**
	Bi	0.180
	W	**0.725**
	Sn	0.154

4.1. Descriptive statistics

The summary of the descriptive statistics for the dataset is presented in Table 3. According to Tompkins et al. (2000), a normally distributed dataset should have a skewness value within the interval of −3 to 3. Based on the findings in Table 2, Mo, Cu, Bi, and Sn have skewness values of 8.245, 5.312, 5.523, and 7.234 respectively. The result obtained shows that variables of Mo, Cu, Bi, and Sn are not normally distributed and volatile which might affect the prediction of the dependent variable (Au). The skewness value obtained for Ni, As, and W falls within normality skewness indicating that Ni, As and W are normally distributed and non-volatile.

4.2. Sensitivity analysis

Sensitivity and correlation analysis were performed to confirm the multicollinearity test. This analysis is shown in Fig. 2 and Table 4. The coefficient of both sensitivity and correlation analysis between the independent and dependent variables were all positive (Table 4). This shows that the independent variable is directly proportional to the dependent variable. The sensitivity and correlation coefficient of Mo, Cu, Bi, and Sn were all less than 0.5 indicating a weak relation between them and the dependent variable. However, the coefficient of sensitivity

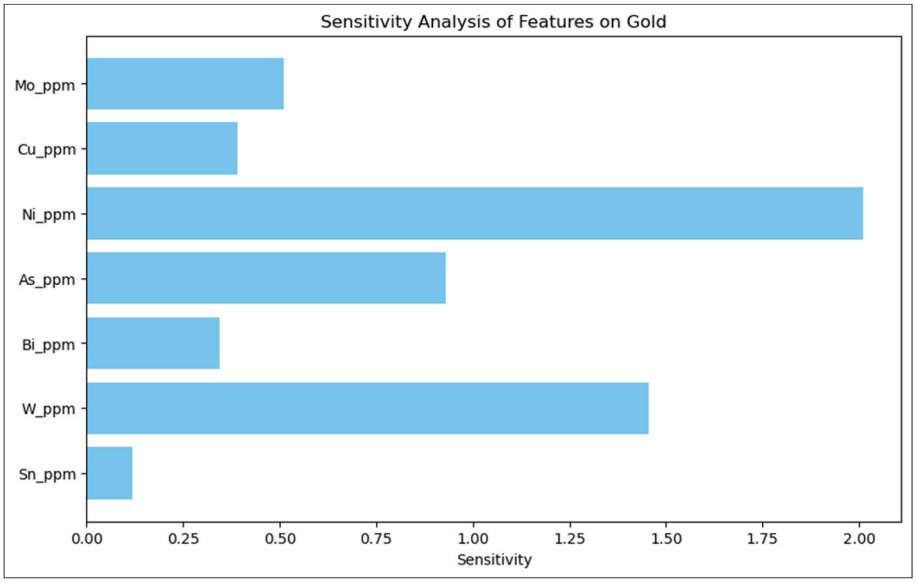

Fig. 2. Sensitivity analysis indicating the influence of the independent parameters on the dependent (Au) occurrence.

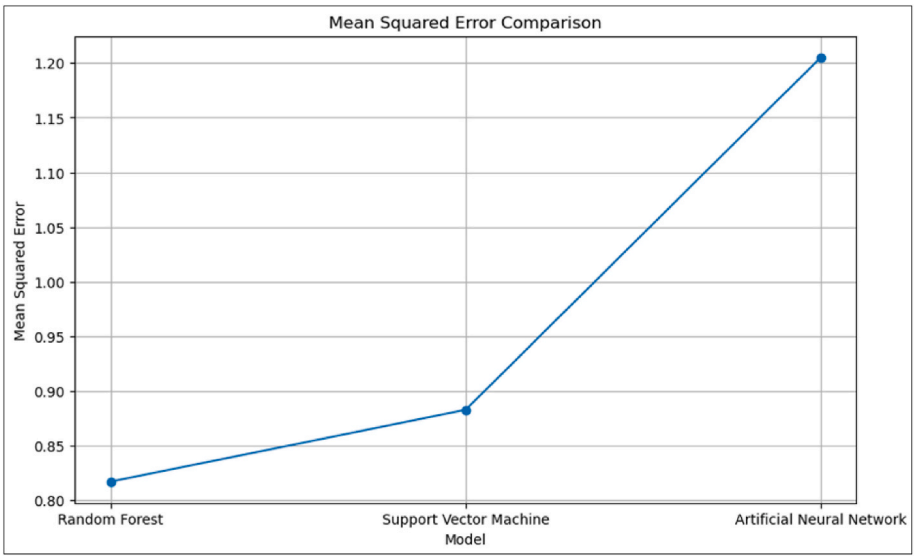

Fig. 3. A comparative plot of mean square error (MSE) variations in Models.

of Ni, As, and W were greater than 0.50 and their correlation coefficient is also greater than 0.50 (Table 4). These indicate that these variables strongly influence the prediction of the dependent variable (Au). The predictor variables of Ni, As, and W were chosen as pathfinder variables for the prediction of Au based on the multicollinearity test, sensitivity, and correlation analysis (Table 4). The three machine learning models (Random Forest, Support Vector Machine, and Artificial Neural Network) were applied for the prediction of Au.

5. Discussion

5.1. Impact of Model configuration on accuracy of predictions

The model configuration in data-driven predictive modeling significantly influences prediction accuracy. In this study, Mean Squared Error (MSE) serves as the metric for assessing training accuracy. Fig. 3 illustrates variations in MSE across three ML models. Overall, Random Forest (RF) and Support Vector Machine (SVM) models demonstrate superior accuracy and stability compared to Artificial Neural Network (ANN) models, yielding MSE values below 0.89 and showing reduced sensitivity to parameter adjustments. The ANN models, especially when using a learning rate above 0.20, exhibit more complex and less precise predictions during training, as indicated by fluctuating MSE patterns.

The number of trees in RF, the width parameter in SVM, and the number of neurons in ANN represent key machine learning (ML) characteristics that define model architecture and influence model complexity. In various scientific fields, it is widely recognized that complex ML models often yield more accurate predictions (Murphy et al. 2012). However, this study specifically examined the MSE outcomes as architecture-related parameters increased in prospective modeling. As depicted in Fig. 3, the findings indicate that more complex model structures do not necessarily lead to improved performance in forecasting for the research area. All statistical metrics, especially the minimum MSE, show an increase with these parameters.

Conversely, the computational burden escalates notably with an increase in the number of trees, neurons, and width parameters. The modeling approaches discussed here often do not necessitate complex ML model structures. This finding is consistent with prior research (Sun et al., 2019), which indicates that more intricate models do not consistently yield more accurate predictions. The limited size of the training datasets, typically encompassing several dozen locations with and without mineral deposits, likely contributes to these results. Overfitting errors may arise from excessive training and complex architectures, but with such small datasets, achieving the necessary precision in model training is feasible.

5.2. Exploration insight from the incorporated machine learning techniques

Data processing and the associated outcome do have a significant effect on the decision to continue exploration activity or abandon the search for a particular mineral commodity. According to Nunoo et al. (2023), the success of exploration depends chiefly on the obtained data obtained, but, most importantly on the type of analytical methods deployed and the interpretation reached. This implies that poorly analyzed data impedes the true reflection of reality. In an attempt to see the true reflection of all forms of data, scientists often adopt various analytical machinery; that ensure optimum accuracy, less error, compare, and one that works best in each scenario (e.g., Abu et al., 2024b; Forson et al., 2024). Thus, adopting different data processing techniques, though this might be time-consuming but potentially reveal relevant information obscured by other methods. For instance, multivariate statistical analysis of similar data from the Central Tanzania craton by Nunoo et al. (2023) shows no form of correlation between gold and other 26 different elements (including well-established pathfinder elements of gold (i.e. As, Cu, Ni, Mo, Bi, W and Sn; e.g., Somarin et al., 2021). Even in the most recent work by Abu et al. (2024a) in a similar terrain shows no form of correlation between gold and notable pathfinders such as Cu, Ni, As and Co among other elements. This observation may possibly arise from a similar multivariate statistical method used. Comparing the data processing techniques from the aforementioned work to the currently incorporated ML techniques (RF, SVM and ANN) suggest the latter being robust in improving the degree of correlation between Au and the selected pathfinders. For example, not only has the latter method shown improved correlative outcome but values multiple times to the usual statistical techniques (e.g., Au and Cu = 0.251 in Table 4, and, Au and Cu = - 0.03 of Table 2 in Nunoo et al. (2023)).

However, an alternative approach via machine learning models and sensitivity analysis (*this study*) revealed a promising correlation between gold and pathfinder of arsenic, tungsten, and nickel (Table 4). Even among the other pathfinder elements of Mo, Bi, Sn, and Cu show weak correlation, such revelation complements the aspect of the geospatial distribution patterns, especially in the southern part of the area compared to the gold anomalous map (Fig. 4). It is thus important to point out that the adopted ML techniques and the sensitivity analysis

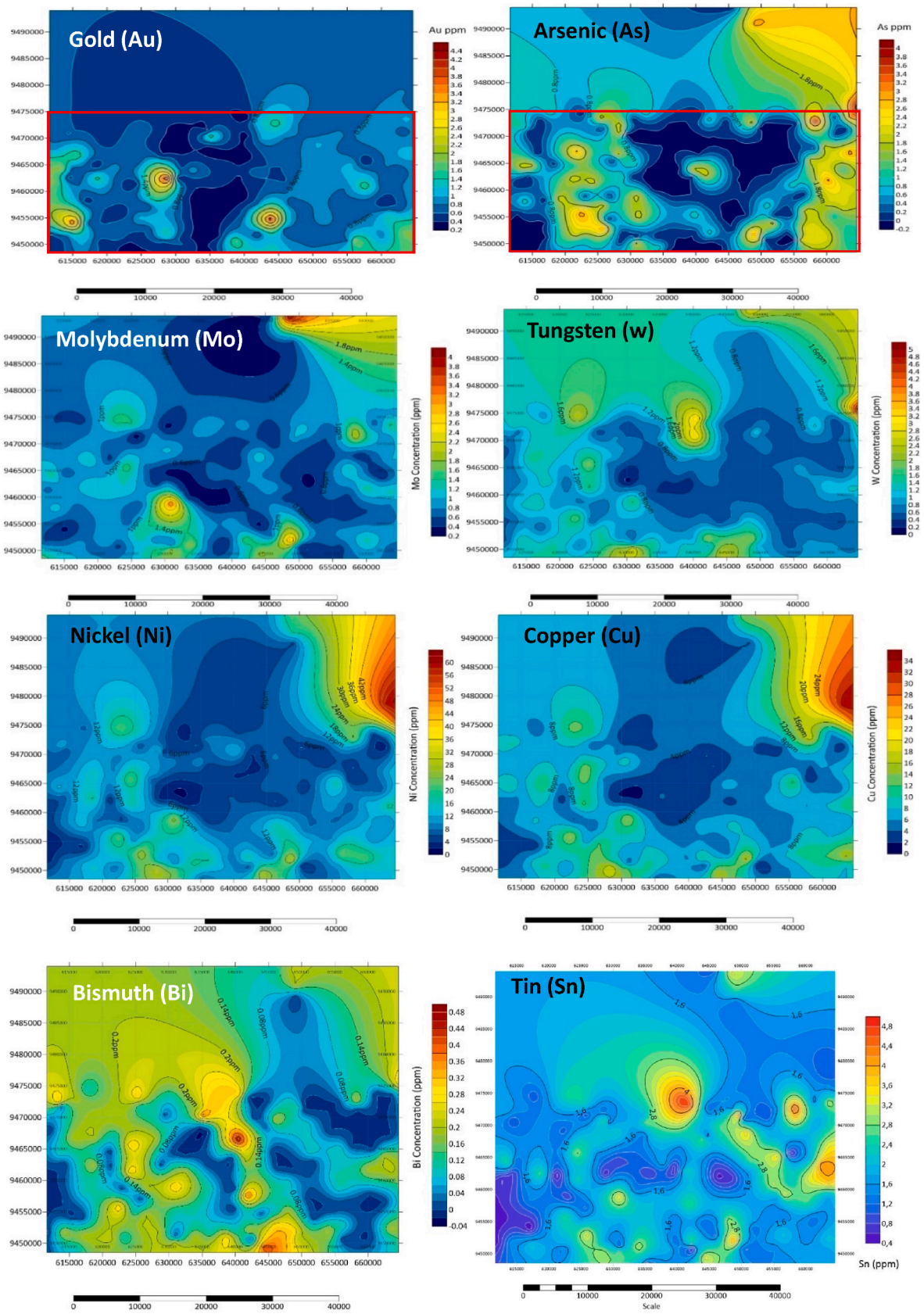

Fig. 4. Geospatial distribution of gold (Au) and selected pathfinder elements (As, Mo, W, Ni, Cu, Bi and Sn) in the area. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

have projected important information masked from multivariate statistical and principal component analysis. The latter techniques are effective but when combined with the machine learning approach and the incorporated sensitivity analysis allow for an optimum understanding of the data.

The sensitivity analysis has unmasked and shown that exploration for gold in the Central Tanzania Craton should pay attention to element of arsenic (As), nickel (Ni), and tungsten (W), as these are potential pathfinders to gold occurrence in this terrain. In addition, the ML techniques of Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) are made of robust mathematical and statistical algorithms or computations depicted by the equation shown earlier, and thus, offer a comparative predictive model plus a measure of validation as mean square error (Fig. 3). The models adopted in this current study, proved effective; but comparatively, the RF was the most efficient ML technique in predicting reliable pathfinder elements in the search for gold mineralization within central parts of the TC. The robustness of the adopted technique is the check for multicollinearity of the predictors (As, Ni, W, Mo, Bi, Sn, and Cu), and their correlation and sensitivity influence of prediction on the dependent variable (Au). This approach highlights the trusted outcome of this method, and our result indicated that these variables showed no multicollinearity and were a strong predictor (especially for As, Ni, and W) for the prediction of Au deposit. The modeling results indicate that the RF model achieves the best prediction performance with an MSE of less than 0.82, followed by the SVM model (0.89), and the last being the ANN model with MSE of 1.21. In contrast, the RF model outperforms the rest of the ML models in overall predictive performance and predictive efficiency. Thus, in the Central parts of the TC, RF is the most preferable ML technique, although the other models of SVM and ANN are of good output considering the data used in this research.

6. Conclusions

The study focused on predictive modeling for mineral prospects in central Tanzania. It involved analyzing 166 stream sediment samples that included gold (Au) values. Evidential characteristics related to gold mineralization were utilized as inputs to train a range of machine learning models, including RF, SVM, and ANN.

Several machine learning models, including ANN, SVM, and RF, were trained on input datasets, and their performance was assessed. Throughout the training process, SVM and RF models exhibited outstanding precision and robustness against variations in model parameters. To determine the optimal model for generating future maps, the predictive capabilities of these trained models were thoroughly evaluated and compared. The Mean Squared Error (MSE) was utilized to evaluate the predictive performance of the machine learning models. Among them, the RF model achieved the lowest MSE of 0.82, followed by the SVM model with an MSE of 0.89, and the ANN model with an MSE of 1.21. These results indicate that the RF model comparatively outperforms the other ML models in terms of predictive accuracy, especially in this study. Based on sensitivity analysis, the most reliable indicators for guiding the search for gold mineralization in the central Tanzanian Craton are tungsten (W), arsenic (As), and nickel (Ni). Sensitivity analysis serves as a valuable tool in machine learning approaches, revealing aspects of data analytics that traditional kriging methods and multivariate statistical approaches may have obscured or presented subtly.

CRediT authorship contribution statement

Samuel Nunoo: Writing – review & editing, Writing – original draft, Supervision, Methodology, Conceptualization. **Mahamuda Abu:** Writing – review & editing, Writing – original draft, Validation, Methodology, Conceptualization. **Emmanuel Ayitey:** Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Conceptualization. **Benatus Norbert Mvile:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Investigation, Data curation. **John Desderius Kalimenze:** Writing – review & editing, Writing – original draft, Project administration, Investigation, Formal analysis, Data curation.

Declaration of competing interest

The author declares that there are no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Abu, M., Mvile, B.N., Kalimenze, J.D., 2024a. Provenance studies of Au-bearing stream sediments and performance assessment of machine learning-based models: insight from whole-rock geochemistry central Tanzania, East Africa. Environ. Earth Sci. 83, 105. https://doi.org/10.1007/s12665-024-11419-2.

Abu, M., Musah, R., Zango, M.S., 2024b. A combination of multivariate statistics and machine learning techniques in groundwater characterization and quality forecasting. Geosystems and Geoenvironment 3, 100261. https://doi.org/10.1016/j.geogeo.2024.100261.

Abu, M., Zango, M.S., Nunoo, S., Anim-Gyampo, M., 2023. Groundwater characterization including prediction of the quality, fluoride, and nitrate occurrence in a typical artisanal mining area in Ghana: a hydrochemical and multivariate statistical approach. Groundwater for Sustainable Development 23, 10104. https://doi.org/10.1016/j.gsd.2023.101041.

Allibone, A.H., McCuaig, T.C., Harris, D., Etheridge, M.A., Munroe, S., Byrne, D., Amanor, J., Gyapong, W., 2002. Structural Controls on Gold Mineralization at the Ashanti Gold Deposit. Society of Economic Geologists, Obuasi, Ghana, p. 65e93. Special Publication 9.

Atanga, F., Amponsah, P.O., Nunoo, S., Kwayisi, D., Forson, E.D., Akabzaa, T.M., Prosper Nude, P.M., 2023. The geology and geochemistry of the Rhyacian Josephine gold deposit, Northwest Ghana. B. Appl. Earth Sci. https://doi.org/10.1080/25726838.2023.2260583.

Bayari, E.E., Foli, S.K., Gawu, Y., 2019. The glacial transport and physical partitioning of mercury and gold in till: implications for mineral exploration with examples from central British Columbia, Canada. Environ. Earth Sci. 78, 268.

Borg, G., Shackleton, R.M., 1997. The Tanzania and NE Zaire cratons. In: DeWit, M., Ashwal, L.D. (Eds.), Greenstone Belts. Oxford University Press, Oxford, pp. 608–619.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Chapman and Hall/CRC, London, UK, p. 368.

Brown, W.M., Gedeon, T.D., Groves, D.I., Barnes, R.G., 2000. Artificial neural networks: a new method for mineral prospectivity mapping. Aust. J. Earth Sci. 47, 757–770.

Chamberlain, C.M., Tosdal, R.M., 2007. U–Pb Geochronology of the Lake Victoria Greenstone Terrane, Tanzania. Mineral Deposit Research Unit the University of British Columbia (Research Program on World-Class Gold Deposits and Advanced Exploration Projects Owned And/or Joint Ventured to Barrick Gold, Placer Dome, Anglogold-Ashanti, Resolute Mining NL as Main Sponsors, 2007.

Cook, Y.A., Sanislav, I.V., Hammerli, J., Blenkinsop, T.G., Dirks, P.H.G., 2015. A primitive mantle source for the Neoarchean mafic rocks from the Tanzania Craton. Geosci. Front. 7, 911–926.

Fedo, C.M., Eriksson, K.A., Krogstad, E.J., 1996. Geochemistry of shales from the archean (3.0 Ga) buhwa greenstone belt, Zimbabwe: implications for provenance and source area weathering. Geochem. Cosmochim. Acta 60, 1751–1763.

Feizi, F., Karbalaei-Ramezanali, A., Tusi, H., 2017. Mineral potential mapping via topsis with hybrid ahp–shannon entropy weighting of evidence: a case study for porphyry-cu, Farmahin area, Markazi province, Iran. Natural Resources Research 26, 553–570.

Forson, E.D., Amponsah, P.O., Wemegah, D.D., Ahwireng, M.D., 2024. Random forest-based mineral prospectivity modelling over the Southern Kibi–Winneba belt of Ghana using geophysical and remote sensing techniques. B. Appl. Earth Sci. 0 (0), 1–16. https://doi.org/10.1177/25726838231225055.

Forson, E.D., Amponsah, P.O., 2023. Mineral prospectivity mapping over the Gomoa area of Ghana's southern Kibi-Winneba belt using support vector machine and Naïve Bayes. J. Afr. Earth Sci. 206, 105024.

Forson, E.D., Wemegah, D.D., Hagan, G.B., 2022. Data-driven multi-index overlay gold prospectivity mapping using geophysical and remote sensing datasets. J. Afr. Earth Sci. 190, 104504.

Henckel, J., Poulsen, K.H., Sharp, T., Spora, P., 2016. Lake victoria goldfields. Episodes Journal of International Geoscience 39 (2), 135–154.

Huang, C., Davis, L.S., Townshend, J.R.G., 2002. An assessment of support vector machines for land cover classification. Int. J. Remote. Sens. 23, 725–749.

Kabete, J.M., Groves, D.I., McNaughton, N.J., Mruma, A.H., 2012. A new tectonic and temporal framework for the Tanzanian Shield: implications for gold metallogeny and undiscovered endowment. Ore Geol. Rev. 48, 88–124.

Kalimenze, J.D., Abu, M., Mvile, B.N., 2023. Soil geochemistry and multivariate statistical assessment of copper–gold-PGEs mineralization in parts of singida region

of the Tanzania craton, Tanzania, East Africa. Arabian J. Geosci. 16, 59. https://doi.org/10.1007/s12517-022-11148-5.

Kwelwa, S.D., Dirks, P.H., Sanislav, I.V., Blenkinsop, T., Kolling, S.L., 2018. Archaean gold mineralization in an extensional setting: the structural history of the kukuluma and matandani deposits, geita greenstone belt, Tanzania. Minerals 8 (4), 171.

Mathew, T.G., Ariffin, K.S., 2018. Gold Potential Mapping in Kelantan (Malaysia) Using Arcgis and Excel Applying Frequency Ratio Model.

McLennan, S.M., Hemming, S., McDaniel, D.K., Hanson, G.N., 1993. Geochemical Approaches to Sedimentation, Provenance and Tectonics, vol. 284. Geological Society of America Special Paper, pp. 21–40.

Mvile, B.N., Abu, M., Kalimenze, J., 2021. Trace elements geochemistry of in situ regolith materials and their implication on gold mineralization and exploration targeting, Dodoma Region, East Africa. Mining, Metallurgy & Exploration 38 (5), 2075–2087.

Nude, P.M., Asigri, J.M., Yidana, S.M., Arhin, E., Foli, G., Kutu, J.M., 2012. J.M. Identifying pathfinder elements for gold in multi-element soil geochemical data from the Wa- Lawra belt, northwest Ghana: a multivariate statistical approach. Int. J. Geosci. 3 (1), 62, 2012.

Nunoo, S., Mvile, B.N., Abu, M., Kelimenze, J.D., 2023. The search for plausible economic mineral deposits in the central parts of Tanzania; insight from stream sediment geochemistry, multivariate statistics and geostatistics. Heliyon 9, e22702. https://doi.org/10.1016/j.heliyon.2023.e22702.

Nzulu, G.K., Högberg, H., Eklund, P., Hultman, L., Nude, P.M., Yaya, A., Magnuson, M., 2023. Pathfinder elements and indicator minerals of Au from the Kubi Gold ore deposits in Ghana. Environ. Earth Sci. 82, 386. https://doi.org/10.1007/s12665-023-11058-z, 2023.

Omali, T.U., 2021. Utilization of remote sensing and gis in geology and mining. Int. J. Sci. Res. in Multidisciplinary Studies 7 (4).

Plouffe, A., 2001. The glacial transport and physical partitioning of mercury and gold in till: implications for mineral exploration with examples from central British Columbia, Canada. Geol Soc Lond 185, 287.

Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M., Chica-Rivas, M., 2015. Machine learning predictive models for mineral prospectivity: an evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol. Rev. 71, 804–818.

Somarin, A., Zhou, L., Steinhage, I., 2021. Application of handheld XRF on Ta-Nb-Sn-W ore: Factory calibration or user calibration? Geochem. J. 55, 149.

Sun, T., Chen, F., Zhong, L., Liu, W., Wang, Y., 2019. Gis-based mineral prospective mapping using machine learning methods: a case study from Tongling ore district, eastern China. Ore Geol. Rev. 109, 26–49.

Yuksel, B., Arica, E., 2018. Assessment of toxic, essential, and other metal levels by ICP MS in lake Eymir and Mogan in Ankara, Turkey: an environmental application, Atom. Spectros 39 (5).

Zhang, N., Zhou, K., Li, D., 2018. Back-propagation neural network and support vector machines for gold mineral prospective mapping in the Hatu region, Xinjiang, China. Earth. Sci. Inform. 11, 553–566.