

Hydrothermal origin of pyrite in the Zawar Pb-Zn deposit, Aravalli Craton: Insights from trace element geochemistry and supervised machine learning algorithms

Sima Gorai*, Bulusu Sreenivas and T. Vijaya Kumar

CSIR-National Geophysical Research Institute, Uppal Road, Hyderabad 500007, India.

Corresponding author: sima.geol90@gmail.com

ABSTRACT

The application of machine learning algorithms on geochemical datasets is emerging as a powerful tool for characterising ore deposits. In this work, we integrate the trace element geochemistry of different generations of pyrite with machine learning algorithms to understand the nature of ore deposition in the Zawar Pb–Zn deposit. Pyrites from Zawar have long been debated to have either sedimentary exhalative (SEDEX) or hydrothermal origins, making them ideal for testing classification approaches. We framed this as a binary classification problem and employed three ML algorithms, Random Forest (RF), Gradient Boosting (GB), and AdaBoost (AB), well-suited for recognising subtle geochemical patterns. A global dataset comprising 727 SEDEX and 577 hydrothermal pyrite samples, characterised by trace elements, such as Co, Ni, Cu, Zn, As, Ag, Sb, and the Co:Ni ratio, provided the training base for these models. Applying the trained models to in-situ Laser Ablation–Inductively Coupled Plasma–Mass Spectrometry (LA-ICP-MS) data from Zawar pyrites, yielded high accuracies: 97.24% for RF, 96.93% for GB, and 96.63% for AB. The classifications overwhelmingly support a hydrothermal origin, for the deposit. These results not only validate earlier geological interpretations but also demonstrate the integrated use of ore-geochemistry on machine learning for understanding the ore-forming processes, ultimately strengthening exploration techniques and mineral deposit models in economic geology

Keywords: Trace elements, Pyrite classification, Hydrothermal processes, Machine Learning, Zawar Pb-Zn deposits

INTRODUCTION

vein-type mineralization. Satellite imaging has since revealed hydrothermal alteration zones along major structures, Machine learning is increasingly applied across diverse branches of Earth Science, including geology, geophysics, hydrology, and atmospheric science, to analyze complex datasets, identify patterns, and improve predictions of natural processes and resource distribution (Gorai et al., 2025a, Dalai and Kumar, 2026). Geochemical exploration of mineral deposits evolved from limited data to handling large datasets, prompting the adoption of machine learning (ML), which has proven effective in classifying ore deposits in the reconnaissance stage of mineral exploration. Our current study uses machine learning to explore the long-debated origin of the Zawar deposit. This area is known for its complex ore formation, with theories pointing to both sedimentary exhalative (SEDEX) and hydrothermal origins. Early work by Ghosh (1957) and Mookherjee (1964) suggested a hydrothermal model supported by structural controls and high mineralization temperatures around 500°C. Later, Talluri et al. (2000) studied fluid inclusions in sphalerite-galena ores, showing temperatures from 320°C (vein-type) to 150°C (massive galena), consistent with hydrothermal deposition. Bhattacharya and Bull (2010) reported mixed features, both SEDEX-type bedding and structurally controlled especially near the Main Zawar Fold (Gorai et al., 2024). In a recent research low temperature hydrothermal Acanthite veins has been reported from the study area (Gorai et al., 2026).

Despite extensive research on the Zawar deposit, a definitive consensus on its genesis remains elusive due to the coexistence of SEDEX and hydrothermal features, leaving a critical

need for objective, data-driven methods to distinguish between these processes across different mineralization stages. In this study, we apply three machine learning models, Random Forest, Gradient Boosting, and AdaBoost, to trace element data (Co, Ni, Cu, Zn, As, Ag, Sb, and Co:Ni ratio) from different pyrite generations: early framboidal pyrite (Py-1), euhedral pyrite (Py-2), and late-stage coarse pyrite (Py-3). These models classify whether each generation formed through SEDEX or hydrothermal processes, based on early published data sets. This approach helps clarify the genesis of the Zawar deposit using data-driven insights from pyrite geochemistry. This method may be useful in mineral exploration as a screening and support tool for interpreting pyrite trace element data from new samples. Once LA-ICP-MS data are obtained, the trained model can help provide a first-order indication of whether the pyrite geochemistry is more consistent with hydrothermal or SEDEX affinity.

GEOLOGICAL SETTING OF THE ZAWAR Pb-Zn DEPOSIT

The Zawar deposit lies within the highly mineralised Proterozoic Aravalli Supergroup in northwestern India (Deb et al., 1989). The Aravalli Craton stretches around 700 km in length along northwest India, is endowed with some of the country's most significant ore deposits, including Pb–Zn mineralisation at Zawar, Rajpura Dariba, and Rampura Agucha; copper deposits at Khetri and Neem Ka Thana; the phosphorite deposits of Jhamarkotra; and the gold deposits of Bhukia and many other economically significant ore deposits (Figure 1). The Aravalli Supergroup is subdivided into Lower-Middle