IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 30, No. 10, October, 2025, pp. 2583-2594
Journal homepage: http://iieta.org/journals/isi

Machine Learning in Geometallurgy: A Review of Advances and Case Studies from Peru's Mining Sector

Martha Ysabel Bedregal Aguilar, Jesus Alberto Torres-Guerra, Luis M. Soto-Juscamayta*

Faculty of Geological, Mining, Metallurgical and Geographic Engineering Universidad Nacional Mayor de San Marcos, Lima 07011, Peru

Corresponding Author Email: luis.soto14@unmsm.edu.pe

https://doi.org/10.18280/isi.301006

Received: 17 September 2025
Revised: 20 October 2025
Accepted: 26 October 2025
Available online: 31 October 2025

Keywords:
geometallurgy, machine learning, predictive modeling, Peru, mineral processing, process optimization

ABSTRACT

The integration of artificial intelligence (AI), particularly machine learning (ML), into geometallurgy provides an important opportunity to optimize mineral processing and mine planning. This study synthesizes recent research on ML-based geometallurgical applications and examines advances and challenges within the Peruvian mining sector. A non-experimental, descriptive methodology was employed through a systematic literature review in Scopus, ScienceDirect, and Web of Science (2013–2023). The search identified 312 records, reduced to 238 after removing duplicates. Following title and abstract screening, 170 studies were excluded, and 33 publications met the inclusion criteria, all reporting ML models incorporated into geometallurgical workflows. The selected studies were classified into six application categories, and two Peruvian case studies were examined: Sociedad Minera Cerro Verde, focused on copper concentration improvement, and the Minsur–Pucamarca Unit, centered on gold leaching optimization. Internationally, research is dominated by supervised classification algorithms for mineralogical prediction, while in Peru successful implementations are mainly associated with computer-assisted decision-making in operational contexts. At Cerro Verde, the use of Random Forest and Gradient Boosting models led to a 6.5% increase in copper production and a 0.8% rise in recovery. At Minsur, the Optimus Leach system improved gold recovery prediction accuracy ($R^2 = 0.81$) and generated USD 1.4 million in economic benefits during its first year. Overall, the findings indicate that ML-enabled geometallurgy can enhance efficiency, profitability, and sustainability when supported by high-quality data, adequate instrumentation, and multidisciplinary teams, contributing to the digital transformation of Peruvian mining.

1. INTRODUCTION

Geometallurgy presents an interdisciplinary approach that links geological variations and variations in mineral processing through block models that include both geological data (mineralogy, lithology, alterations, rock strength, etc.) and process data (recoveries, concentrate grade, reagent consumption, energy consumption, etc.). In this way, it predicts the process response based on the properties of the process feed and its location in the deposit [1].

A geometallurgical program contains a large amount of data of different types from several areas such as geology, mine, concentration plant, leaching, electrodeposition or refining according to the operations of each company and even environment and safety, which allows to have a better knowledge of the mineral deposit and achieve greater efficiency of the process by optimizing the tonnage, grade and recovery of the metal, reducing the environmental impact, providing greater confidence to investors, etc. Therefore, there is a useful database for decision-making in the planning of the production and mining process [2]. Figure 1 shows the flow diagram of a geometallurgical program.

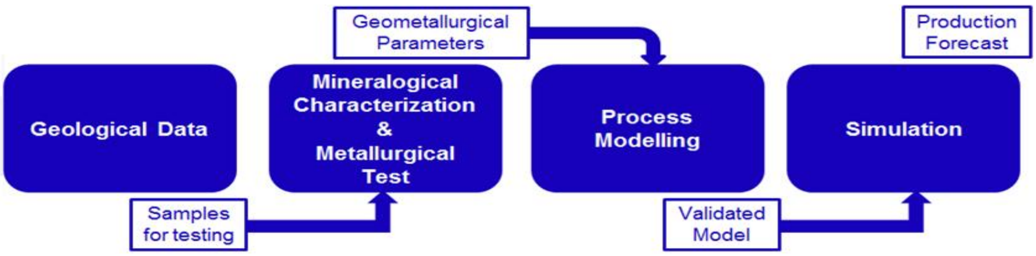


Figure 1. Geometallurgical program focus

Data science and programming have been used in mineral processing since 1970 for modeling, simulation, control and optimization through the implementation of instrumentation, expert systems, reinforcement learning algorithms and neural networks, since the nineties a branch of artificial intelligence called machine learning has taken relevance to be used as a control system and decision support [3]. Artificial intelligence and machine learning applied to process engineering has had cycles of enthusiasm in research and disappointment due to the lack of significant impacts, however with the arrival of the fourth industrial revolution, the internet of things and big data, computational capacity, the variety of data and the improvement in algorithms have increased, which has allowed the success of many applications based on data and / or images [3].

Machine learning predicts trends and has as its main characteristic that the algorithms learn from themselves over time, improving their accuracy without needing to be reprogrammed. It finds patterns by studying a training data set and develops an algorithm without human intervention. However, for its correct operation it depends on the quality of the feed data, properly calibrated instrumentation and training in the various scenarios.

Machine learning algorithms can be integrated into real-time systems, evaluating changes in the process and responding with alternatives to obtain better results. They are used in the prediction of missing data, forecasting impact parameters, validating algorithms using visualization and statistical tools such as accuracy, cluster distortion, confusion matrix, receiver operating characteristic curve, squared error, which allow the evaluation of predicted data; in addition, they can be coupled to optimization methods, increasing their exploration capacity and identifying data that are more significant in a faster way, allowing a greater number of simulation runs [4].

There are three types of algorithms: supervised learning based on labeled data, unsupervised learning based on unlabeled data, and reinforcement learning based on reward or penalty. Machine learning modeling methods have faster, multidimensional processing capabilities that have not yet been fully utilized to integrate all process properties into the geometallurgical block model [5]. Figure 2 presents the main machine learning algorithms.

The objectives of this study are to describe the main applications of machine learning in the field of geometallurgy found in research in recent years, to present successful cases of this type of application in the main mining units in Peru that allow understanding the current situation of the Peruvian mining sector and to identify barriers in the implementation of machine learning applications.

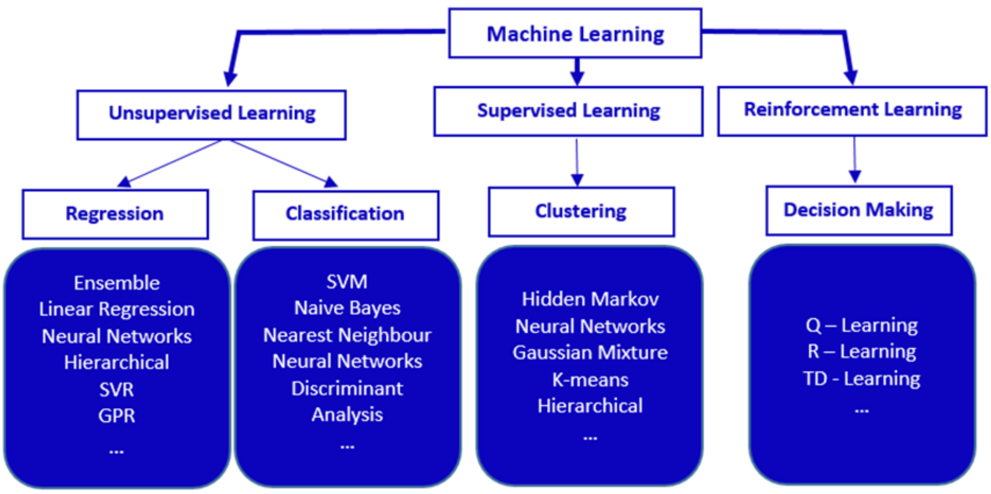


Figure 2. Types of machine learning techniques

2. MATERIALS AND METHODS

This study adopts a descriptive, non-experimental research design aimed at identifying, classifying, and analyzing the main applications of machine learning in the field of geometallurgy, with special emphasis on the advances and challenges of the Peruvian mining sector. The methodological process was structured in four main stages: literature review, data classification, case study selection, and synthesis-analysis.

2.1 Literature review strategy

A comprehensive search was conducted for publications published between January 2013 and December 2023 in high-impact scientific databases such as Scopus, ScienceDirect, and Web of Science. Search strings were formulated using a combination of controlled vocabulary and Boolean operators, including:

- "geometallurgy" AND "machine learning"
- "geometallurgy" AND "artificial intelligence"
- "geometallurgy" AND ("predictive modeling" OR "advanced analytics")

Publications were included that:

(a) explicitly address machine learning applications within geometallurgical workflows,
(b) present a methodological description and performance metrics, and
(c) correspond to peer-reviewed articles, conference presentations, or technical reports from recognized events (e.g., PERUMIN).

Works that were excluded were:

(a) did not detail the methodology used,
(b) were exclusively conceptual with no applied results, or
(c) were not related to mining or mineral processing.

2.2 Data extraction and categorization

From each selected publication, relevant information was extracted on the following topics: type of algorithm used, data source and scale, commodity analyzed, process stage, and reported results. The applications were classified into six categories, following the framework proposed by Koch and Rosenkranz [2] and updated with recent contributions [4]:

(a) mineralogical prediction models,
(b) data-driven geometallurgical modeling,
(c) yield prediction for mine planning,
(d) computer-aided decision-making in operations,
(e) environmental impact prediction, and
(f) data-centric laboratories.

2.3 Selection of case studies in Peru

Two Peruvian mining operations were selected that document the successful use of machine learning in geometallurgical processes:

•Cerro Verde Mining Company (optimization of copper concentration).

•Minsur – Pucamarca Unit (optimization of gold leaching).

The selection criteria were: (a) availability of verifiable quantitative results, (b) detailed public presentation of the methodology in technical forums such as PERUMIN 35/36, and (c) representativeness in terms of mineral type and operational scale.

2.4 Analytical approach

The analysis integrated a descriptive synthesis of global trends with a comparative evaluation of the Peruvian cases. Each application was evaluated considering:

•Type and architecture of machine learning algorithms implemented (e.g., Random Forest, Gradient Boosting, SVM, ANN).

•Data integration strategy (real-time sensors, laboratory data, historical operating databases).

•Performance indicators measured (recovery, throughput, energy efficiency, cost reduction).

•Barriers and enabling factors for implementation.

The synthesis sought to contrast international scientific production with the Peruvian operational context, highlighting both the technological potential and the implementation challenges towards the transition to digital mining.

3. RESULTS

In the literature review it has been found that the main applications of machine learning in geometallurgy currently developed can be classified as shown in Figure 3 in: (a) mineralogical prediction models, (b) data-driven geometallurgical modeling, (c) performance prediction for mine planning, (d) computer-aided decision making in operations, (e) prediction of environmental impacts and (f) data-centric laboratories, with a greater number of investigations in mineralogy prediction.

Figure 3 illustrates the global distribution of machine-learning applications in geometallurgy, with mineralogical prediction and spatial modeling dominating the literature. This concentration of studies reflects the availability of large image-based datasets (SEM, hyperspectral, optical microscopy), which are well-suited for supervised algorithms such as Random Forest or CNNs. Conversely, categories such as environmental prediction or digital-twin integration remain underrepresented due to limited access to long-term environmental datasets and the technological gap between universities and full-scale mining operations. The disparate distribution indicates that while the foundational components of digital geometallurgy are well-developed, downstream applications involving operations control or sustainability are still emerging.

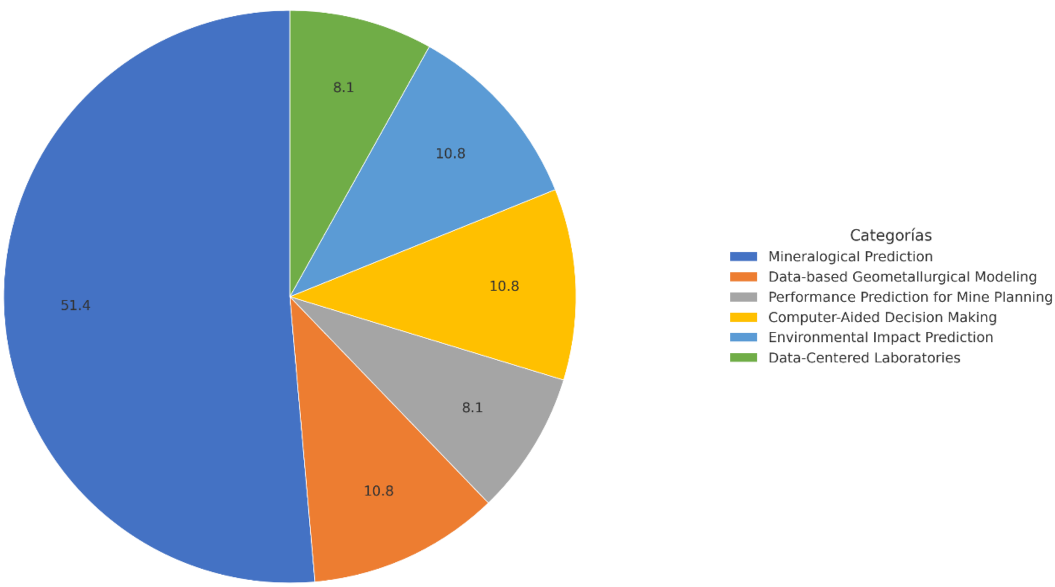


Figure 3. Publications on machine learning applications in the field of Geometallurgy

3.1 Mineralogical prediction

Among the objectives of geometallurgical studies is the maximization of the information extracted from mineral deposits. Thus, through the modeling and simulation of concentration processes, a correlation is made between textural classes (grain size, shape, and mineral associations) and their behavior in the comminution and/or flotation process to predict metallurgical results (recovery, grade, particle distribution). Therefore, quantified textural information is a relevant indicator for mining planning and process optimization [6].

Mineral liberation analysis is important for the control and optimization of comminution circuits and is used in geometallurgy to understand the effect of particle behavior on mineral processing. 1D and 2D release spectra can be obtained in scanning electron microscopes (SEM) or optical microscopes, while 3D release analysis requires an X-ray computed tomograph. Currently, more economical methods are being developed to evaluate the degree of liberation through optical micrographs using supervised image classification machine learning algorithms such as Random Forest Tree, which uses the statistical properties of minerals by evaluating mineral colors and textures to subject them to a voting process and automatically classify particle images, allowing to distinguish metallic minerals from gangue. In this way, 2D mineralogical maps can be estimated, and the variation in the release spectra, mineralogical composition, and the degree recovery curve can be measured [7, 8].

Machine learning algorithms and descriptors are also being used to develop automated mineralogical characterization methods of drill cores for the estimation of modal mineralogy and textural classification, which reduces the time and cost of these tests that were traditionally performed in geological logging with the qualitative description of lithology, mineralogy, mineral texture and element analysis with X-ray fluorescence (XRF). Among the descriptors for the extraction of image features are: (a) the gray level co-occurrence matrix (GLCM) that evaluates pixel frequency and (b) local binary patterns (LBP) that evaluate geometric patterns. As for the machine learning algorithms, Random Forest trees are used, based on the voting of many decision trees; Support Vector Machine (SVM), based on a hyperplane that maximizes the distance between characteristic vectors belonging to different classes, and artificial neural networks (ANN), based on the construction of a function represented as a series of weighted sums that are organized in layers to minimize the classification error [9].

Recent studies have established this category as one of the most advanced areas within digital geometallurgy, particularly due to the increasing use of hyperspectral imaging, automated microscopy, and supervised machine learning models. For example, the systematic review by Jung and Choi [10] highlights that algorithms such as Random Forest, SVM, and neural networks are the most frequently used for mineral texture classification and modal mineralogy estimation; however, the authors warn that training quality depends heavily on image standardization and spectral homogeneity, which remains a challenge in deposits with high lithological variability. The study by Tuşa et al. [11] demonstrated that combining hyperspectral data with high-resolution mineralogical information improves the prediction of mineral abundances using RF, SVM, and ANN; however, they identified the spatial integration of images, spectra, and SEM-MLA data as the main limitation, as it can introduce errors in large drilling campaigns. Finally, a recent deep-learning study applied to drill-core images showed that convolutional neural networks with transfer learning can predict mineral content directly from RGB images, although their performance is still constrained by the limited amount of training samples and the availability of labeled datasets [12]. Taken together, these works demonstrate significant progress in automated mineralogical analysis, yet they also raise debate over the representativeness of 2D models compared to 3D approaches, where the higher accuracy obtained through tomography comes with substantially higher operational costs.

3.2 Data-driven geometallurgical modeling

For the spatial modeling of process properties, regression models, multivariate statistics such as principal component analysis and partial least squares, and geostatistical methods such as krigging are generally used; however, machine learning methods can also be used, which, while requiring a large amount of training data, are fast to process and can handle multidimensionality. For this purpose, in the research by Lishchuk et al. [5], they evaluated ten machine learning algorithms and compared them considering the relative standard deviation, obtaining that the decision tree algorithms are the most appropriate for modeling non-additive variables such as recovery.

An alternative to generating geometallurgical models is the use of supervised learning and linear regression in the prediction of metallurgical variables, which has the advantage of reducing the time and costs invested in laboratory tests. For example, in the research by Mu and Salas [13] for the construction of the geometallurgical model of a copper deposit they carried out the identification of domains evaluating four methods Kmeans, hierarchical clustering (AGG), spatial clustering of applications with noise based on density (DBSCAN) and self-organization maps (SOM), obtaining better results in Kmeans under the Silhouette and Calinsky Harabasz cluster validity indices, in terms of dimensionality reduction they found that unsupervised neural networks called autoencoder work better than the principal component analysis technique (PCA), in block modeling they used the computational geometric technique Alpha Shape for the identification of blocks without information and subsequently performed the interpolation with the supervised learning regression method Gradient Boosting. Concluding that it is feasible to generate a geometallurgical model based only on the data methodology.

Data-driven geometallurgical models powered by machine learning have gained relevance due to their ability to capture nonlinear relationships between plant variables and geological characteristics. The study by Goronovski et al. [14] compared ten algorithms and identified decision trees as the most suitable for modeling recoveries, due to their flexibility when dealing with non-additive variables. However, these models face limitations in their extrapolation capacity: they perform well within the training range, but their performance drops when encountering new ore types or operational conditions not previously seen.

Other studies such as those by Hasan et al. [15] combine dimensionality reduction techniques (autoencoders) with Gradient Boosting to generate complete spatial models, reporting improvements over traditional PCA. However, the use of autoencoders introduces algorithmic opacity, making interpretation difficult for geometallurgical personnel.

3.3 Performance prediction for mine planning

The purpose of mining extraction sequence planning is to achieve greater profitability in the long-term mining operation. To achieve this, sequential decisions are made in the extraction of minerals based on the maximization of reward functions that take into account weighted values of economic, productive, and environmental factors, and the minimization of regret functions that consider probable losses. In this sense, numerous process simulation tools can be used, including band algorithms, which work well with geological variables characterized by uncertainty [2].

Prediction models for key mineral processing performance indicators can be developed using multiple linear regression machine learning algorithms and neural networks with independent variables such as rock characteristics linked to plant-dependent variables such as mill tonnage, recovery, reagent and ball consumption, taking into account mining fleet management records. These prediction models, integrated with simultaneous stochastic optimization models, allow for the generation of mining sequencing that generates greater profitability for the company due to a better evaluation of the block value of the geometallurgical model, taking into account the associated energy costs for grinding and reagent consumption, as well as more precise compliance with production schedules [16, 17].

Both global research trends coincide in noting that integrating mine sequencing, metallurgical variables and ML improves profitability [18] demonstrated that integrating predictive models with stochastic algorithms enables better estimation of higher-value blocks. The main contribution is the ability to incorporate geological and metallurgical uncertainty, something classical deterministic models cannot achieve.

However, debate persists regarding the stability of these predictions: some authors note that, under operational fluctuations, models may amplify variability, affecting day-to-day planning. In addition, building robust models requires complete datasets that many operations do not possess [19].

3.4 Computer-assisted decision-making in operations

Modeling in mining aims to improve production planning through the characterization of geometallurgical units and process optimization, taking into account high-impact parameters such as tonnage, recovery, and grade. To achieve this, machine learning has proven to be an extremely useful tool that requires the knowledge and experience of operations experts and a basic understanding of algorithms and data science.

The models used can be divided into three types: (a) Physical models derived from first principles which require a full understanding of the phenomenon and all parameters must be measured, (b) Phenomenological models which have measured physical limits and hard-to-measure parameters are calibrated with empirical constants requiring experimental design and pilot testing and (c) Data-driven models where inputs are derived from their correlations and distribution are developed from historical data, they are accurate within training data useful in brownfield operations. Machine learning regression models are data-driven models that are developed from historical sensor data and metallurgical chemical assays, so they are being used in thin section microscopy, flotation, comminution, hydrocyclone evaluation, etc. [20].

Two control strategies have been developed in mineral processing, which are expert systems and predictive models, and currently hybrid models are emerging that integrate simple prediction models into expert systems, however the efficiency of these systems depends on the quality of the data provided by the instruments, among which stand out near infrared spectroscopy (NIR), Fourier transform spectroscopy (FTIR) and Raman spectroscopy, virtual sensors based on spectroscopy and microelectromechanics can also be used, which has the advantage of overcoming the aggressive chemical conditions of the environment [21].

This category includes the most applied studies [22] showed that the use of AutoML in flotation and comminution can reduce calibration times and improve predictive accuracy compared to traditional linear models. The main contribution lies in automating the model-selection process; however, its limitation is that AutoML tends to select complex models requiring costly computational infrastructure.

Globally, there is consensus that hybrid models (expert systems + ML) offer greater stability; however, some studies warn that the quality of recommendations depends almost entirely on sensor calibration and the consistency of historical data recurring weakness in real mining operations [23].

3.5 Prediction of environmental impacts

The sustainability, competitiveness and success of the mining industry in the era of digital transformation known as the fourth industrial revolution depend on the development of energy-efficient and environmentally conscious methods for mineral processing, such as the optimization of selective mining units to reduce gangue extraction; non-explosive rock breaking methods that reduce the generation of fine material and promote microfracturing of minerals; use of tunnel boring machines for thin ore bodies; pre-concentration of minerals through automated mineral sorting with sensors based on optical microscopy technologies and image recognition based on artificial intelligence; optimization of mineral concentration by analyzing flotation froths with machine learning tools; circular economy strategies that involve the reuse of metallurgical waste; remote monitoring of mining operations; reduction of carbon-based energy use; and ecosystem restoration. The combination of these new methods will reduce the environmental impact of mining [24].

Life cycle assessment is a tool for measuring the environmental impacts generated by mining operations, taking into account crucial indicators such as global warming potential (GWP), terrestrial acidification (TA), water depletion (WD), and land use (LU). This tool is of great importance today due to the increasing demand for metals due to the global transition to clean energy and, consequently, the increase in production. Life cycle analysis integrated with geometallurgy, process simulation, and machine learning algorithms such as decision trees, neural networks, and Random Forests achieves more accurate and reliable results. In this way, the fate of contaminants can be predicted and, therefore, production processes can be designed and/or modified with a circular economy approach [25].

The prediction of acid rock drainage allows for the assessment of environmental risk from pre-feasibility stages and is a support for the valorization of geo-environmental model blocks due to the economic costs involved in the mine closure stage; to define geo-environmental models, the evaluation of the mineralogical, textural, geochemical, geometallurgical properties of the rock, and microbiological processes is required. Currently, field tests (drill core observations, leaching tests for pH and dissolved metals, total metal concentration analysis with portable X-ray fluorescence equipment), static tests (acid-base accounting (ABA) tests, net acid generation (NAG) tests, acid buffer characteristic curve method), and kinetic tests (laboratory leaching column tests, cell moisture tests, and field pad tests) are being carried out to classify waste according to its acid-generating potential. However, due to the multiple processes that are involved in the formation of rock acid drainage, more accurate predictions are required that can be achieved using machine learning algorithms [26]. Machine learning algorithms have a significant impact on GHG emissions reduction in LCA for differentiated geographic approaches; however, they are still underutilized [27].

Recent research combines LCA with decision trees, neural networks, and Random Forest to estimate emissions and acid drainage. The study by Stehlik et al. [28] reported that ML models reduce prediction errors of GWP and TA compared to traditional methods but also highlight that the availability of environmental databases remains limited, reducing model generalization.

Regarding acid drainage, the study by Anthony et al. [29] describe that ML algorithms can integrate mineralogical, textural, and geochemical properties; however, they warn that microbial variability and hydrogeological conditions are still not adequately modeled. This raises an ongoing controversy: to what extent can models be trusted when part of the underlying phenomenology is not yet represented?

3.6 Data-centric laboratories

The mining value chain generates a large amount of data in each of its processes. This has been particularly true in the area of mineral processing, where the flow of data has increased due to the implementation of sensors. Historically, this data has been processed separately. However, the current trend is toward geometallurgy and the joint processing of data, as this allows for greater efficiency in mineral extraction. This is why data-centric laboratories are emerging, whose primary objective is data generation and experimentation. This requires instrumentation and infrastructure for real-time monitoring, analysis, and characterization of the operation, as well as data analytics methods. In dry laboratories, data management, multidisciplinary data analysis, research, experimentation, process design and training are carried out. For this purpose, they use data science, statistics, simulation that may include virtual and augmented reality, machine learning and visualization tools. Unlike wet or conventional laboratories, they do not have materials, instruments and reagents but they do require a much larger investment due to the essential components for the operation of this type of laboratories such as: (a) technology that involves high-performance computing equipment, data storage servers and communication and information infrastructure, (b) integration of knowledge since data from tests and sensors will be used as data from automatic analysis of the process, (c) expert personnel in operations as well as data scientists, (d) data governance to ensure that data management meets accessibility, security and quality standards, (e) operational and financial models that ensure the viability and sustainability of this type of project. Finally, the objective is to have better mineral processing evaluation indicators that allow us to obtain greater efficiency in the operation and cost reduction [30].

For the capture of relevant quality information, instrumentation and equipment are required in the main mining operations, for example in explorations the following can be used: (a) high resolution remote sensing such as the hyperspectral imaging method that allows the mapping of geological characteristics based on optical reflectance properties, (b) drone-based sensors that identify mineral anomalies to determine mineral exploration targets, (c) portable measuring devices such as laser-induced breakdown spectroscopy (LIBS), portable X-ray fluorescence (pXRF) for geochemical analysis in rocks, (d) cross-sectional seismic tomography to determine geological formations; in mining there is rock glass technology that can predict geological and geotechnical conditions prior to mining. There are also detection methods such as the electromagnetic spectrum and ground penetrating radar, as well as automated mechanical rock cutting, rock preconditioning investigations with water jets, thermal impulses or explosive impulses. In terms of mineral processing, there are (a) automated sampling, (b) particle image analyzers in wet, dry, dynamic flows, etc., (c) fully integrated automation (TIA) such as digital twins that are presented as a virtual simulation of mineral processing operations [21].

Studies by Yao et al. [31] indicate that data-centric laboratories represent the next technological leap, enabling integration of sensors, digital twins, and advanced analytics for virtual experimentation. Their major contribution is the ability to accelerate experimentation cycles without operating physical equipment.

The main limitation is economic: the necessary infrastructure (HPC, servers, OT/IT integration, data governance) exceeds the capacity of many mid-sized operations. In addition, there is ongoing controversy regarding dependence on external vendors and who ultimately controls data and model ownership [32].

3.7 Success stories in Peru

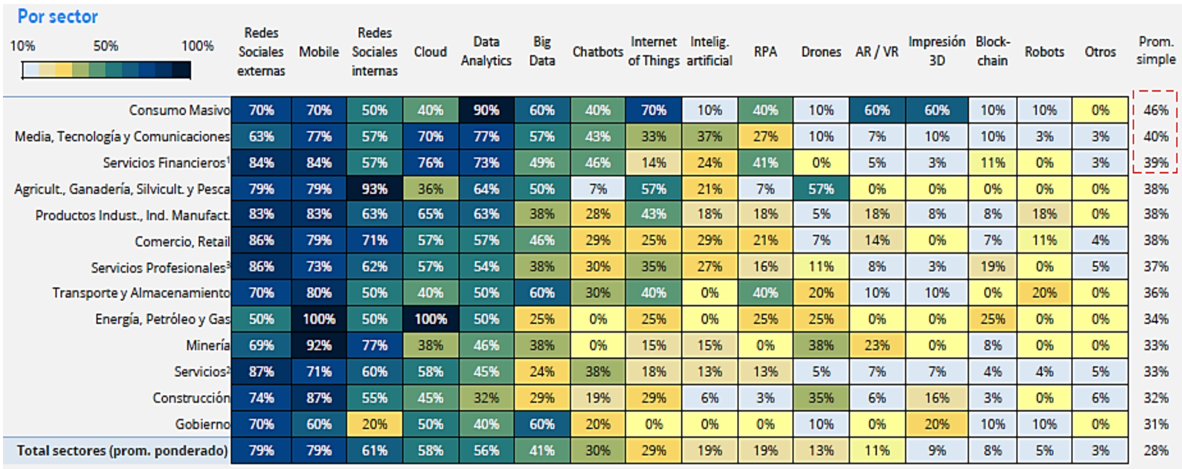
According to the Global AI Adoption Index 2023, 42% of companies worldwide have adopted artificial intelligence, while in Latin America, an increase from 40% to 47% in AI adoption has been reported by 2023. The main limitation is the lack of experience or knowledge in the field, followed by ethical concerns regarding data governance. Investment is focused on process automation, and the main reason for adoption is achieving faster decision-making to improve customer experience.

In Peru, 44% of companies are undergoing a digital transformation process, with the most advanced sectors being mass consumption, communications, and financial services. The main problems they face are a lack of human resources, budget, and strategies. They have high data availability, but it is underutilized. The mining sector has been working on digital transformation for an average of three years, and as shown in Figure 4, only 15% of companies are at a mature stage and have artificial intelligence applications.

As shown in Figure 4, the sharp growth in machine-learning publications after 2018 is consistent with the expansion of open-access mineralogical databases, cloud-based computing resources, and the democratization of Python-based data-science tools. The increase also coincides with declining ore grades worldwide, which has pressured mining companies to invest in predictive modeling of recovery and throughput. This pattern suggests that the growth is not merely academic but also driven by operational needs in real mining environments.

Figure 5 shows the number of publications on machine learning applications by category of the main Peruvian mining companies presented at Perumin 35 and 36, noting that there is an increase in publications related to the category of computer-assisted decision-making.

Figure 5 shows an evolution in the categories of machine-learning-related publications presented at PERUMIN between 2013 and 2023. The shift toward ‘geometallurgy’, ‘digital operations’ and ‘process optimization’ after 2018 reflects the accelerated adoption of digital transformation initiatives in Peruvian mining. This trend coincides with an increase in sensorization of concentrator plants, wider availability of automated mineralogical data (SEM-MLA, hyperspectral scanning), and industry pressure to optimize recovery amid lower ore grades. In contrast, earlier conferences were dominated by topics such as blasting or geomechanics, which require less high-density data. The growing prominence of AI-enabled studies at PERUMIN suggests that Peruvian operations are moving from descriptive to predictive and prescriptive analytics, aligning with international trends and indicating a maturing of data infrastructures within major mining companies.



Por sector (10% – 50% – 100%)	Redes Sociales externas	Mobile	Redes Sociales internas	Cloud	Data Analytics	Big Data	Chatbots	Internet of Things	Intelig. artificial	RPA	Drones	AR / VR	Impresión 3D	Block-chain	Robots	Otros	Prom. simple
Consumo Masivo	70%	70%	50%	40%	90%	60%	40%	70%	10%	40%	10%	60%	60%	10%	10%	0%	46%
Media, Tecnología y Comunicaciones	63%	77%	57%	70%	77%	57%	43%	33%	37%	27%	10%	7%	10%	10%	3%	3%	40%
Servicios Financieros[1]	84%	84%	57%	76%	73%	49%	46%	14%	24%	41%	0%	5%	3%	11%	0%	3%	39%
Agricult., Ganadería, Silvicult. y Pesca	79%	79%	93%	36%	64%	50%	7%	57%	21%	7%	57%	0%	0%	0%	0%	0%	38%
Productos Indust., Ind. Manufact.	83%	83%	63%	65%	63%	38%	28%	43%	18%	18%	5%	18%	8%	8%	18%	0%	38%
Comercio, Retail	86%	79%	71%	57%	57%	46%	29%	25%	29%	21%	7%	14%	0%	7%	11%	4%	38%
Servicios Profesionales[3]	86%	73%	62%	57%	54%	38%	30%	35%	27%	16%	11%	8%	3%	19%	0%	5%	37%
Transporte y Almacenamiento	70%	80%	50%	40%	50%	60%	30%	40%	0%	40%	20%	10%	10%	0%	20%	0%	36%
Energía, Petróleo y Gas	50%	100%	50%	100%	50%	25%	0%	25%	0%	25%	25%	0%	0%	25%	0%	0%	34%
Minería	69%	92%	77%	38%	46%	38%	0%	15%	15%	0%	38%	23%	0%	8%	0%	0%	33%
Servicios[2]	87%	71%	60%	58%	45%	24%	38%	18%	13%	13%	5%	7%	7%	4%	4%	5%	33%
Construcción	74%	87%	55%	45%	32%	29%	19%	29%	6%	3%	35%	6%	16%	3%	0%	6%	32%
Gobierno	70%	60%	20%	50%	40%	60%	20%	0%	0%	0%	10%	0%	20%	10%	10%	0%	31%
Total sectores (prom. ponderado)	79%	79%	61%	58%	56%	41%	30%	29%	19%	19%	13%	11%	9%	8%	5%	3%	28%

Figure 4. Digital transformation technologies

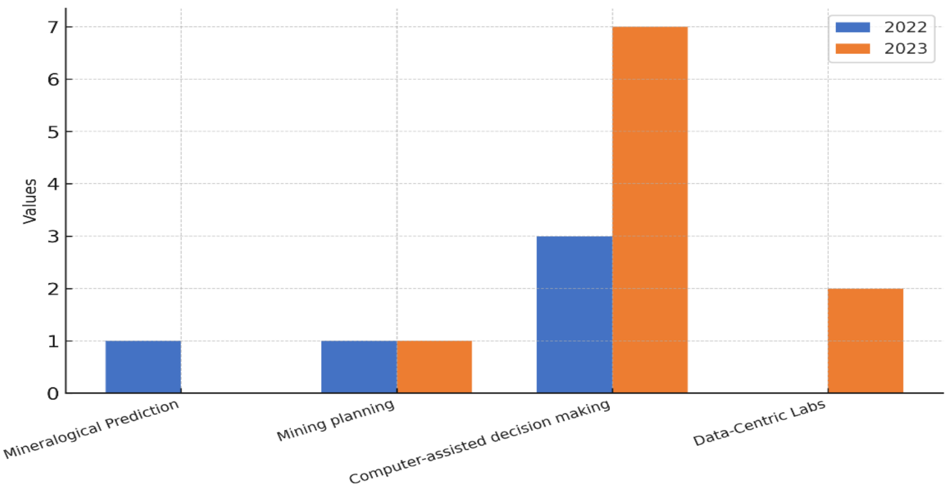


Figure 5. Publications on machine learning applications in the field of Geometallurgy in Perumin 35 and 36

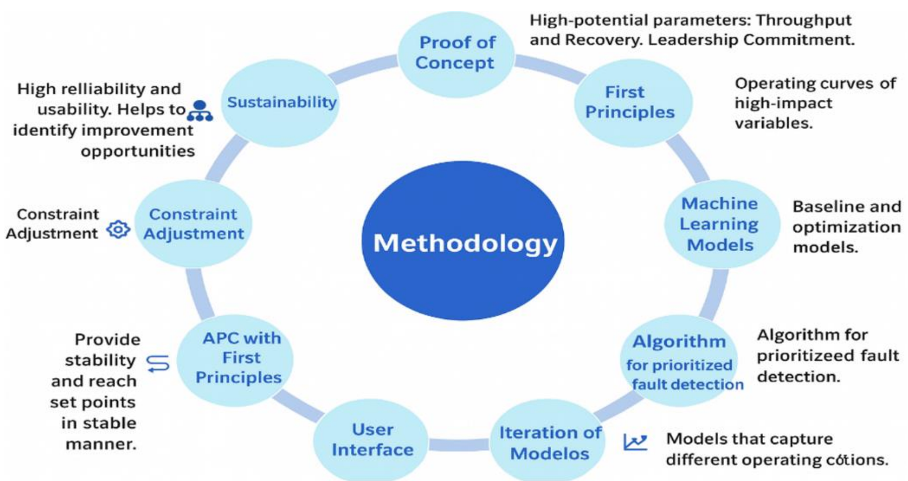


Figure 6. SMCV Artificial Intelligence Implementation Methodology

Cerro Verde Case. In recent years, Sociedad Minera Cerro Verde has been promoting the application of artificial intelligence in its operations. The study [33] in his research work presented at Perumin 35 explains the success story of increasing the production of pounds of copper concentrate in Cerro Verde, as can be seen in Figure 6, the applied methodology has nine phases.

Figure 6 contrasts global and Peruvian publication patterns, revealing that Peru exhibits a disproportionate focus on operational applications, such as plant control and metallurgical optimization, rather than early-stage research. This bias reflects Peru's status as a copper-dominated producer where concentrator plants already have extensive instrumentation, allowing rapid adoption of ML for predictive control. Conversely, upstream research—such as advanced mineral-texture analysis—is less represented due to the limited presence of specialized research facilities in the region.

•Proof of concept defining high-potential variables such as recovery and tonnage, which are correlated with ore type, Cu and Fe head grade, hardness, oxidation level, and operating parameters such as cyclone pressure, % mill discharge solids, % rougher feed solids, rougher concentrate grade, scavenger concentrate grade, primary collector dosage, secondary collector dosage, and pH.

•Implementation of first principles, which are the ranges established for the models to ensure metallurgically consistent recommendations.

•Construction of both baseline and optimization recovery and tonnage models, useful for comparing production with and without AI. The machine learning algorithms used were Random Forests and, specifically, gradient boosting, which consist of thousands of decision trees that allow for multiple scenarios and nonlinear relationships to be presented, helping the models generate better recommendations for optimizing processes.

•Development of the anomaly detection module that allows for the evaluation of data consistency and, therefore, the validation of the quality of the model input information, which comes from permanently monitored plant instrumentation.

•Model iteration or model training, which is very important due to variations in mineral type, especially when mining phase changes, equipment changes, or flowsheet modifications occur. In all these cases, model retraining is required.

•User interface, where operators view recommendations and have the option to approve or reject them, following discussions involving operators and supervisors.

•The importance of aligning the APCs with first principles to enable the development of artificial intelligence and ensure the stable execution of recommendations.

•Adjusting constraints, which are basically the maximum design capacities of the equipment.

•Sustainability, which involves building trust among end users, namely, operators, so they will use the system permanently.

Among the conclusions, it is noted that the implementation of AI allowed for an increase in copper pounds by 6.5% and copper recovery by 0.8%. The correct functioning of the APCs, automatic control, and instrumentation were of utmost importance for this work from a technical point of view, and from a human resources point of view, the leadership of the management, communication between the various areas involved, and training of operating personnel. Furthermore, this tool has allowed for the identification of several opportunities for improvement, since the artificial intelligence recommendations motivate the evaluation, interpretation, and discussion of these among metallurgists.

In the case of Sociedad Minera Cerro Verde, the reviewed technical documentation indicates that the predictive models were trained using operational variables such as feed grade (CuT and CuS), mineralogical composition from MLA analyses, hardness indices (Axb, BWi), particle size distributions (P80), flotation air flow, reagent dosage and plant throughput. Before model training, datasets were cleaned through outlier removal based on interquartile ranges and z-score thresholds, followed by normalization of continuous variables. Feature importance analyses showed that mineralogy, feed grade and grind size had the highest predictive weight. Model validation was performed using 10-fold cross-validation and independent test sets, reporting performance metrics such as R^2, RMSE and MAE. Despite the positive results, internal reports highlight limitations including sensor calibration drift, missing data during plant upsets, and model sensitivity to mineralogical variability not previously observed in the training dataset.

Minsur Case. Optimus Leach is the real-time recommendation system for optimizing the gold leaching process at the Pucamarca mining unit, presented as a success story at Perumin 36. The main objective is to maximize the gold recovery percentage from the leaching cells. The methodology includes the following steps:

•Data Collection and Preparation: Mineralogical variables such as alterations, ore grades, and fine ore content were taken into account, as well as process variables such as particle size distribution, pH, moisture content, irrigated cells, irrigation rates, irrigation ratios, cyanide concentration, and gold and silver grades. These variables were obtained from the following data sources: (a) Stockpile Control, (b) Drip Irrigation Cells, (c) Daily Report, (d) Cell Certificate, (e) Moisture Content, (f) Mining Plan, and (g) ADR Parameters.

•Algorithm Selection: For deterministic models, the Klimpel equation was considered, while for assembled machine learning models, the type of algorithm used was not specified. It should be noted that recommendation systems typically use Random Forest (RF), Gradient Boosting (GB), Support Vector Machines (SVM), and Neural Networks (ANN).

•Development of Predictive Models: For the development of this system, a model for estimating daily percolated ounces and a model for estimating daily percolated flow were used. Both were built using machine learning algorithms based on a deterministic model that allows theoretically estimating the grade based on irrigation days in the cell and statistical models that explain gold contribution based on operational variables.

•Model Training: The performance of the models was evaluated by partitioning the data, i.e., one data set for pattern training and another data set for model validation testing.

•Real-time implementation: The web application has three parts: (a) Diagnostics, where irrigation rates and historical recoveries can be viewed; (b) Recommendations for cell irrigation rate, duration, and intensity parameters for optimization; and (c) Simulation, where different cell parameter and gold recovery scenarios can be created. The architecture for this case can be seen in Figure 7.

•Results evaluation: Figure 8 shows that the base deterministic model has a correlation of 0.6 and an error of 0.16, while the machine learning model has a correlation of 0.81 and an error of 0.05, thus reducing the variability in gold grade estimation in the rich solution, which meant greater recovery and more efficient management of leach pad inventories.

•Impact: An economic benefit of $1.4M was achieved in the first year. At the organizational level, planning was decentralized and the team was empowered with machine learning, in addition to the implementation of enablers such as data integration.

Figures 7 and 8 summarize the economic and operational impacts of ML-based systems implemented in Peru. The improvements observed—such as increased recovery and more stable reagent consumption—reflect the ability of ML models to capture nonlinearities that conventional linear control strategies cannot. However, the results also highlight the dependency on high-quality sensor data, as shown by performance drops during periods of incomplete or noisy measurements. These figures underscore both the potential of ML for process intensification and the infrastructural challenges that must be addressed for long-term adoption.

For the Minsur–Pucamarca operation, the Optimus Leach system integrates variables such as cyanide concentration, pH, dissolved oxygen, solution flow rate, ore permeability, granulometry, mineralogical composition and heap height. Preprocessing steps include smoothing of noisy time-series signals, interpolation of incomplete data and normalization of chemical parameters. The Gradient Boosting model used in Optimus Leach was validated using k-fold cross-validation and evaluated with metrics such as R^2 and RMSE, yielding prediction accuracies consistent with industrial requirements. However, the system documentation indicates that its performance may decrease when abrupt changes in ore permeability occur or when heap irrigation is irregular, which introduces temporal variability that the model does not fully capture. Additionally, the limited availability of high-frequency mineralogical data constrains the representation of mineralogical heterogeneity within the heap.

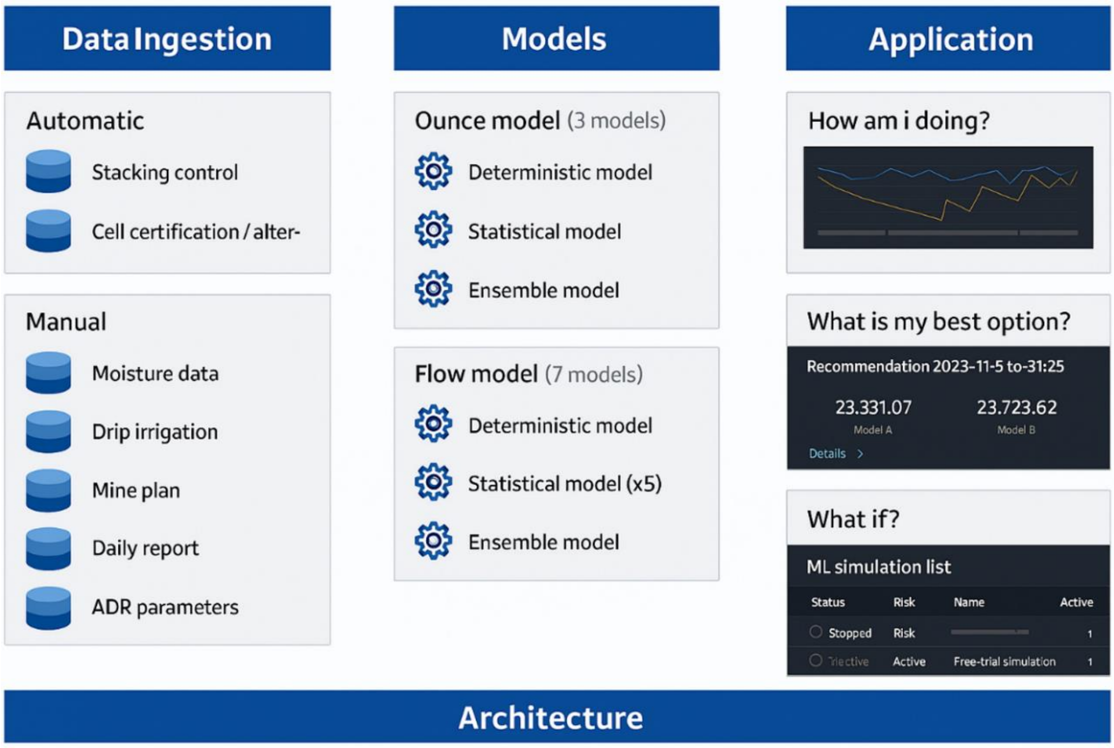


Figure 7. Optimus leach Minsur application architecture

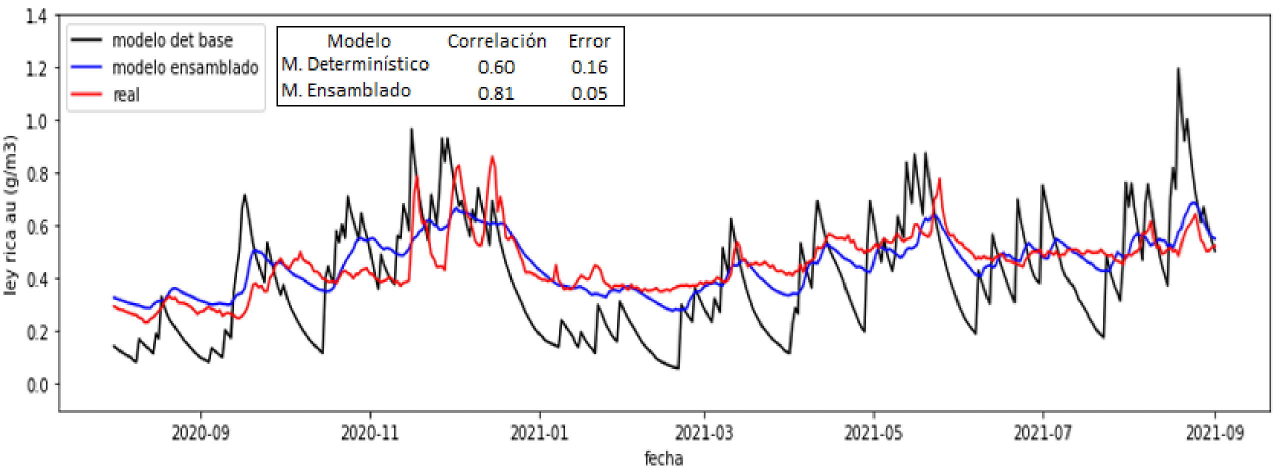


Figure 8. Comparison of real curve with deterministic model and machine learning model

4. CONCLUSIONS

In recent years, international research on machine learning applications in geometallurgy has focused on mineralogical prediction models, data-driven geometallurgical modeling, yield prediction for mine planning, computer-aided decision-making in operations, environmental impact prediction, and data-centric laboratories. However, this focus is notably on mineralogical prediction using supervised classification algorithms.

This growing emphasis illustrates a broader theoretical shift from deterministic geometallurgical workflows toward hybrid computational frameworks, where machine learning models enhance —rather than replace— geological and metallurgical interpretation. This integration reveals the need for more robust data architectures, higher levels of standardization, and the incorporation of explainable ML techniques to ensure interpretability and adoption within operational teams.

In the Peruvian mining sector, only 15% of companies have artificial intelligence applications, and they are successfully adopting applications focused on computer-aided decision-making in the operations area, as evidenced by the cases presented by Cerro Verde and Minsur with their machine learning-based recommendation systems, achieving increases in tonnage and recovery in their operations, which ultimately translates into greater profitability. It is important to highlight that the main mining companies in Peru are pursuing advanced analytics strategies with a view to becoming digital mines in the near future, characterized by integrated operation centers, the implementation of artificial intelligence in all their processes, and high levels of automation with machine learning predictive models coupled with expert systems.

These cases demonstrate the practical implications of ML deployment, showing that predictive modeling can stabilize metallurgical performance, anticipate operational deviations, and support decision-making in near-real time. As Peruvian operations continue moving toward digital transformation, machine learning becomes a key enabler for semi-autonomous concentrator plants and integrated remote operation centers.

The greatest challenge in the development of machine learning-based recommendation systems has been the formation of multidisciplinary teams with experience in operations, as well as knowledge of data science. To quickly overcome this challenge, mining companies have opted for joint ventures with startups that have data scientists and analytical scientists. It should also be considered that the databases of the different mining areas are not interconnected and often require large investments in instrumentation and telecommunications networks for data capture, as well as rigorous instrument maintenance and calibration programs.

This reveals a structural limitation that must be addressed: the historical fragmentation of operational databases. Future implementations will require unified data governance strategies, sensor calibration protocols, and interoperability frameworks that guarantee data quality, continuity and traceability—conditions without which advanced ML models cannot achieve stable performance.

They claim that a geometallurgical program allows for greater efficiency in operational processes. On the other hand, machine learning algorithms allow predictions to be made through the interpretation of data and patterns connected with established knowledge [4]. In accordance with the bibliographic review and the applications presented, it is observed that geometallurgy, by integrating the disciplines of geology, mining, and metallurgy, has led to a better understanding of the entire value chain, providing a tool for better estimating the mineral values of the blocks. This allows the development of models with more precise maximization or minimization objective functions and the establishment of action strategies for process optimization. Machine learning algorithms also complement geometallurgy and are strengthened for the construction of efficient models that allow improving planning and production in mining [2].

Given these synergies, future work should explore the integration of 2D and 3D mineralogical datasets into unified ML workflows, develop interpretable models to support domain experts, and incorporate uncertainty quantification to enhance block valuation and risk assessment.

For machine learning algorithms to function correctly, quality input data, instrumentation, and training are required. Similarly, the case analysis shows that data understanding, traceability, and preparation are very important for successful predictions with machine learning models. This is usually the longest process in the execution of this type of project, since the main mining operation areas traditionally have isolated databases that need to be interconnected to identify patterns and trends. Investments in instrumentation, sensors, and telecommunications networks are often also required to capture the data that feeds the model in real time.

Mineral processing control strategies are classified into expert systems, predictive models, and hybrid models, which are the integration of the two previously mentioned [4]. Accordingly, Peruvian mining companies that are using some artificial intelligence application have had initial success with hybrid models related to core business variables such as throughput and recovery executed in expert systems. Therefore, the next stage could be to develop applications that can include other variables such as mill power, which is related to energy consumption, which would have, in addition to economic benefits, a reduction in the environmental impact and carbon footprint.

Consequently, lines of future research should focus on expanding ML applications toward sustainability metrics, including predictive models for acid drainage, water balance optimization, and carbon emission reduction. Additionally, adaptive control systems that directly link ML outputs to operational actuators represent a promising frontier, enabling dynamic, real-time optimization of metallurgical performance.

ACKNOWLEDGMENT

Thanks to the professors of the Postgraduate Unit of the Faculty of Geological, Mining, Metallurgical and Geographical Engineering of the National University of San Marcos, especially Dr. Jesús Alberto Torres Guerra for his teachings and encouragement to publish this article.

REFERENCES

[1] Lishchuk, V., Koch, P.H., Ghorbani, Y., Butcher, A.R. (2020). Towards integrated geometallurgical approach: Critical review of current practices and future trends. Minerals Engineering, 145: 106072. https://doi.org/10.1016/j.mineng.2019.106072

[2] Koch, P.H., Rosenkranz, J. (2020). Sequential decision-making in mining and processing based on

geometallurgical inputs. Minerals Engineering, 149: 106262. https://doi.org/10.1016/J.MINENG.2020.106262
[3] McCoy, J.T., Auret, L. (2019). Machine learning applications in minerals processing: A review. Minerals Engineering, 132: 95-109. https://doi.org/10.1016/J.MINENG.2018.12.004
[4] Ghoroghi, A., Rezgui, Y., Petri, I., Beach, T. (2022). Advances in application of machine learning to life cycle assessment: A literature review. The International Journal of Life Cycle Assessment, 27(3): 433-456. https://doi.org/10.1007/S11367-022-02030-3/TABLES/1
[5] Lishchuk, V., Lund, C., Ghorbani, Y. (2019). Evaluation and comparison of different machine-learning methods to integrate sparse process data into a spatial model in geometallurgy. Minerals Engineering, 134: 156-165. https://doi.org/10.1016/J.MINENG.2019.01.032
[6] Lamberg, P. (2011). Particles - The bridge between geology and metallurgy. https://urn.kb.se/resolve?urn=urn:nbn:se:ltu:diva-40210, accessed on Aug. 06, 2025.
[7] Camalan, M., Çavur, M., Hoşten, Ç. (2017). Assessment of chromite liberation spectrum on microscopic images by means of a supervised image classification. Powder Technology, 322: 214-225. https://doi.org/10.1016/J.POWTEC.2017.08.063
[8] Lotter, N.O., Kormos, L.J., Oliveira, J., Fragomeni, D., Whiteman, E.J.M.E. (2011). Modern process mineralogy: Two case studies. Minerals Engineering, 24(7): 638-650. https://doi.org/10.1016/j.mineng.2011.02.017
[9] Koch, P.H., Lund, C., Rosenkranz, J. (2019). Automated drill core mineralogical characterization method for texture classification and modal mineralogy estimation for geometallurgy. Minerals Engineering, 136: 99-109. https://doi.org/10.1016/J.MINENG.2019.03.008
[10] Jung, D., Choi, Y. (2021). Systematic review of machine learning applications in mining: Exploration, exploitation, and reclamation. Minerals, 11(2): 148. https://doi.org/10.3390/MIN11020148
[11] Tuşa, L., Khodadadzadeh, M., Contreras, C., Rafiezadeh Shahi, K., Fuchs, M., Gloaguen, R., Gutzmer, J. (2020). Drill-core mineral abundance estimation using hyperspectral and high-resolution mineralogical data. Remote Sensing, 12(7): 1218. https://doi.org/10.3390/RS12071218
[12] Boiger, R., Churakov, S.V., Ballester Llagaria, I., Kosakowski, G., Wüst, R., Prasianakis, N.I. (2024). Direct mineral content prediction from drill core images via transfer learning. Swiss Journal of Geosciences, 117(1): 8. https://doi.org/10.1186/s00015-024-00458-3
[13] Mu, Y., Salas, J.C. (2023). Data-driven synthesis of a geometallurgical model for a copper deposit. Processes, 11(6): 1775. https://doi.org/10.3390/PR11061775
[14] Goronovski, A., Vind, J., Vassiliadou, V., Panias, D., Tkaczyk, A.H. (2019). Radiological assessment of the Bayer process. Minerals Engineering, 137: 250-258. https://doi.org/10.1016/J.MINENG.2019.04.016
[15] Hasan, A.F., Jabbar, S.F., Raheem, F.S. (2025). Comparative analysis of four programming languages for machine learning. Ingénierie des Systèmes d'Information, 30(6): 1437-1445. https://doi.org/10.18280/isi.300603
[16] Both, C., Dimitrakopoulos, R. (2022). Geometallurgical prediction models of processing plant indicators for stochastic mine production scheduling. IFAC-PapersOnLine, 55(21): 162-167. https://doi.org/10.1016/J.IFACOL.2022.09.261
[17] Boisvert, J.B., Rossi, M.E., Ehrig, K., Deutsch, C.V. (2013). Geometallurgical modeling at Olympic dam mine, South Australia. Mathematical Geosciences, 45(8): 901-925. https://doi.org/10.1007/S11004-013-9462-5
[18] Both, C., Dimitrakopoulos, R. (2021). Applied machine learning for geometallurgical throughput prediction—A case study using production data at the tropicana gold mining complex. Minerals, 11(11): 1257. https://doi.org/10.3390/MIN11111257
[19] Bagger, N.C.F., Sørensen, M., Stidsen, T.R. (2019). Dantzig–Wolfe decomposition of the daily course pattern formulation for curriculum-based course timetabling. European Journal of Operational Research, 272(2): 430-446. https://doi.org/10.1016/J.EJOR.2018.06.042
[20] Koh, E.J., Amini, E., Gaur, S., Maquieira, M.B., Heck, C.J., McLachlan, G.J., Beaton, N. (2022). An automated machine learning (AutoML) approach to regression models in minerals processing with case studies of developing industrial comminution and flotation models. Minerals Engineering, 189: 107886. https://doi.org/10.1016/J.MINENG.2022.107886
[21] Ghorbani, Y., Zhang, S.E., Nwaila, G.T., Bourdeau, J.E., et al. (2023). Dry laboratories–Mapping the required instrumentation and infrastructure for online monitoring, analysis, and characterization in the mineral industry. Minerals Engineering, 191: 107971. https://doi.org/10.1016/J.MINENG.2022.107971
[22] Hu, H., Dai, L., Sun, W., Zhuang, Y., et al. (2022). Some remarks on the electrical conductivity of hydrous silicate minerals in the Earth crust, upper mantle and subduction zone at high temperatures and high pressures. Minerals, 12(2): 161. https://doi.org/10.3390/MIN12020161
[23] Deng, Z., Oraby, E.A., Eksteen, J.J. (2020). Sulfide precipitation of copper from alkaline glycine-cyanide solutions: Precipitate characterisation. Minerals Engineering, 145: 106102. https://doi.org/10.1016/J.MINENG.2019.106102
[24] Nwaila, G.T., Frimmel, H.E., Zhang, S.E., Bourdeau, J.E., Tolmay, L.C., Durrheim, R.J., Ghorbani, Y. (2022). The minerals industry in the era of digital transition: An energy-efficient and environmentally conscious approach. Resources Policy, 78: 102851. https://doi.org/10.1016/J.RESOURPOL.2022.102851
[25] Rachid, S., Yassine, T.A.H.A., Benzaazoua, M. (2023). Environmental evaluation of metals and minerals production based on a life cycle assessment approach: A systematic review. Minerals Engineering, 198: 108076. https://doi.org/10.1016/J.MINENG.2023.108076
[26] Parbhakar-Fox, A., Lottermoser, B.G. (2015). A critical review of acid rock drainage prediction methods and practices. Minerals Engineering, 82: 107-124. https://doi.org/10.1016/J.MINENG.2015.03.015
[27] Milojevic-Dupont, N., Creutzig, F. (2021). Machine learning for geographically differentiated climate change mitigation in urban areas. Sustainable Cities and Society, 64: 102526. https://doi.org/10.1016/J.SCS.2020.102526
[28] Stehlik, P., Smith, R., Varbanov, P.S. (2023). Jiří Jaromír Klemeš–His life and legacy. Journal of Cleaner

Production, 405: 137014. https://doi.org/10.1016/J.JCLEPRO.2023.137014
[29] Anthony, E.J., Marriner, N., Morhange, C. (2014). Human influence and the changing geomorphology of Mediterranean deltas and coasts over the last 6000 years: From progradation to destruction phase? Earth-Science Reviews, 139: 336-361. https://doi.org/10.1016/J.EARSCIREV.2014.10.003
[30] Ghorbani, Y., Zhang, S.E., Nwaila, G.T., Bourdeau, J.E. (2022). Framework components for data-centric dry laboratories in the minerals industry: A path to science-and-technology-led innovation. The Extractive Industries and Society, 10: 101089. https://doi.org/10.1016/J.EXIS.2022.101089
[31] Yao, G., Guo, Q., Li, Y., Xu, Z., Han, Z., He, M., Qi, T. (2022). An innovation technology for recovering silver and valuable metals from hazardous zinc leaching residue through direct reduction. Minerals Engineering, 188: 107857. https://doi.org/10.1016/J.MINENG.2022.107857
[32] Jareño, F., González, M.D.L.O., López, R., Ramos, A.R. (2021). Cryptocurrencies and oil price shocks: A NARDL analysis in the COVID-19 pandemic. Resources Policy, 74: 102281. https://doi.org/10.1016/J.RESOURPOL.2021.102281
[33] Kozhirbayev, Z. (2024). Enhancing neural machine translation with fine-tuned mBART50 Pre-Trained model: An examination with low-resource translation pairs. Ingénierie des Systèmes d'Information, 29(3): 831-838. https://doi.org/10.18280/isi.290304

NOMENCLATURE

AI	Artificial intelligence
ML	Machine learning
ANN	Artificial neural networks
SVM	Support Vector Machine
RF	Random Forest
GB	Gradient Boosting
GWP	Global warming potential
NIR	Near-Infrared spectroscopy