# CHARACTERIZING ORE PARTICLE SIZE FOR SAG MILL FEED CONTROL

by

Matias Saavedra Miranda



A Thesis Submitted to the Faculty of the

DEPARTMENT OF MINING & GEOLOGICAL ENGINEERING

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

2025

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Master's Committee, we certify that we have read the thesis prepared by Matias Saavedra, titled *CHARACTERIZING ORE PARTICLE SIZE FOR SAG MILL FEED CONTROL* and recommend that it be accepted as fulfilling the dissertation requirement for the Master's Degree.

Nathalie Risso
*Nathalie Risso* Date: 06/16/2025

Moe Momayez
*Moe Momayez* Date: 06/16/2025

Victor Octavio Tenorio (Jun 17, 2025 10:29 PDT)
*Victor Tenorio* Date: 06/17/2025

Final approval and acceptance of this thesis is contingent upon the candidate's submission of the final copies of the thesis to the Graduate College.

I hereby certify that I have read this thesis prepared under my direction and recommend that it be accepted as fulfilling the Master's requirement.

Nathalie Risso
*Nathalie Risso* Date: 06/16/2025
Master's Thesis Committee Chair
*Mining and Geological/Geophysical Department*

## DEDICATION

*To myself, I am grateful for the resilience and determination that carried me through this journey. Despite the challenges and uncertainties, I stayed focused on my goals and trusted in my abilities. This achievement reflects not only academic success but personal growth and perseverance.*

*To my beloved wife, thank you for your unwavering support and love. Your courage in leaving home and standing beside me through every step has been my greatest strength. This accomplishment is as much yours as it is mine—a symbol of our shared sacrifices and dreams.*

*To my parents and my brother, though far in distance, your love and values have always guided me. Thank you for your constant support and belief in me. This milestone is a tribute to the foundation you gave me and the dreams we continue to pursue together.*

*To my advisor and the University of Arizona, thank you for the mentorship, support, and opportunities that shaped my academic path. I am truly grateful for the experiences and knowledge gained here, and I carry them forward into the next chapter with pride and appreciation.*

## TABLE OF CONTENTS

# LIST OF FIGURES

**Appendix A – Blending Characterization for Effective Management in Mining Operations**



**Appendix B – Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed**

# LIST OF TABLES

**Appendix A – Blending Characterization for Effective Management in Mining Operations**



**Appendix B – Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed**

## ABSTRACT

This thesis presents a comprehensive, data-driven study for predicting the particle size distribution (PSD) at the feed of a semi-autogenous grinding (SAG) mill. A consistent PSD is essential for efficient energy use, stable throughput, and effective downstream processing. However, feed variability, caused by ore heterogeneity, stockpile segregation, and reactive blending, makes prediction challenging.

The first part of this work reviews the current state of research and industrial practice, synthesizing insights from over 45 peer-reviewed studies on ore blending, stockpile management, and the application of machine learning in mineral processing. The review identifies a gap between the growing availability of high-frequency sensor data and its limited use in real-time predictive tools, highlighting the need for systems that forecast PSD and support data-informed operational understanding.

Building on these insights, the second part develops and validates a machine learning framework using two years of high-resolution operational data from a copper mining operation. The methodology integrates unsupervised clustering with Random Forest regression to forecast key PSD metrics (F10–F90 and TOPSIZE) based on variables such as feeder speeds, stockpile levels, and ore throughput. Cluster-specific models capture nonlinear, regime-dependent behaviors with high accuracy ($R^2 > 0.90$).

In addition to prediction, the thesis presents a data-driven sensitivity analysis to evaluate how changes in input variables, such as individual feeder rates, influence PSD outcomes. This analysis shows that coarser PSD metrics, such as F70 and TOPSIZE, are more responsive to input changes than finer ones. A variability analysis further demonstrates the model's ability to quantify and explain fluctuations in PSD, with predicted distributions exhibiting up to a 15

All components are integrated into a web-based application that allows users to explore model behavior, visualize PSD forecasts, and assess the impact of operating scenarios in real time.

This work advances the field by demonstrating how machine learning and clustering can be effectively applied to model and interpret SAG mill feed variability. The resulting framework offers a practical approach for enhancing predictive insight in mineral processing operations..

# Chapter 1

# Introduction

## A. Motivation

Blending and particle size distribution (PSD) control are central to the stability and efficiency of mineral processing operations. In particular, the performance of semi-autogenous grinding (SAG) mills is highly sensitive to variability in feed characteristics, including ore hardness, moisture content, and PSD. Irregularities in feed composition can lead to frequent disturbances in mill load and power draw, causing suboptimal grinding conditions, accelerated wear on liners, and fluctuations in throughput [1].

A primary contributor to feed variability is the stockpile and feeder system, which plays a more dynamic role than merely buffering mine output. Stockpiles also function as blending and surge control units that integrate materials from different mine faces or process streams. However, they can introduce new challenges such as particle segregation, channeling, and dead zones—leading to inconsistent feed quality at the mill [2, 3]. These effects propagate downstream, disrupting stable operation, reducing energy efficiency, and increasing operational costs.

Energy consumption in grinding processes represents a significant portion of a mine's total energy usage, often exceeding 40% in concentrator circuits [4]. Variability in feed properties can amplify energy demands, impairing mill efficiency and driving up costs. A stable PSD at the SAG mill feed is thus essential not only for optimizing comminution efficiency but also for reducing overall energy consumption.

As the mining industry advances toward more sustainable and intelligent operations, there is growing emphasis on minimizing energy usage, reducing feed variability, and maximizing resource efficiency. Improving the understanding and control of feed PSD, particularly

through enhanced blending strategies, real-time monitoring, and predictive tools, offers a direct path to achieving these goals. Tackling feed variability at its source, namely within the stockpile and feeder system, is a critical step toward more stable, efficient, and sustainable grinding operations.

### B. Problem Statement

Despite the critical role of particle size distribution (PSD) in determining SAG mill performance and downstream efficiency, current industry practices lack the tools needed to anticipate and manage this variability effectively. Although many operations have implemented sensors and automation infrastructure, reliable real-time tools to estimate PSD before milling remain absent. Conventional laboratory-based PSD measurements are delayed and infrequent, limiting their utility for timely decision-making [5]. As a result, operators are often forced to respond to process disturbances only after they have impacted throughput or stability, rather than anticipating and mitigating them in advance [6].

This reactive control strategy leads to unstable mill loads, fluctuations in energy consumption, and wear on critical components such as liners and grinding media. The challenge is compounded by the complex and variable nature of stockpile behavior, which introduces inconsistencies in ore properties,particularly PSD, density, and moisture—despite being designed to buffer upstream variability. While stockpiles are central to blending, their dynamic response is not well understood or effectively modeled in most operational environments.

Although some correlations between operational variables—such as feeder speeds, stockpile levels, and throughput, and PSD have been observed, these insights remain largely qualitative or site specific. There is currently no standardized framework that integrates such multivariate operational data into a predictive model capable of generalizing across different feed conditions [7]. Moreover, existing control systems do not provide actionable guidance to operators on how to adjust inputs to maintain a desired PSD profile, especially when feed characteristics deviate from historical trends. This thesis addresses these gaps by investigating the following core questions:

- Can upstream variables, such as feeder speeds, stockpile levels, and ore flow, be quantitatively linked to the PSD of ore feeding the SAG mill?
- Can machine learning models or data-driven techniques be used to estimate PSD in real time?
- How much could this help improve feed consistency and process stability?

To answer these, the study develops a framework that combines unsupervised clustering with Random Forest regression to predict PSD based on upstream sensor data. It further introduces a controllability analysis to evaluate the extent to which operational setpoints can be adjusted to meet PSD targets. Together, these contributions aim to enhance real-time process control, reduce energy consumption, and improve the sustainability of mineral processing circuits.

### C. Proposal

This thesis proposes the development of a data-driven case of study to predict the particle size distribution (PSD) of the ore feed in real time, using operational data from the stockpile and feeder system. By analyzing correlations between feeder speeds, stockpile level indicators, material throughput, and other process variables, the study applies machine learning models, particularly cluster-guided Random Forest regressor, to anticipate PSD fluctuations before they impact SAG mill performance.

In addition to real-time prediction, the framework includes a controllability assessment that quantifies the system's ability to achieve specific PSD targets within realistic operational conditions. This is achieved through univariate sensitivity analysis and Monte Carlo simulations that evaluate the impact of individual manipulated variables on PSD outputs. Results are integrated into a web-based application, providing operators with an intuitive interface to simulate operational scenarios, visualize PSD outcomes, and assess the feasibility of control strategies.

By linking predictive modeling with actionable insights, this framework supports more consistent feed quality, improved mill efficiency, and reduced energy consumption. Ultimately,

the work contributes to more stable, cost-effective, and sustainable grinding operations in the mining industry.

## D. Goals

### *A. Primary Objective*

- Assess the relationship between ore particle size distribution (PSD) and key operational variables, including feeder speed, level sensor data, and other potential factors, to support consistent feed quality to the SAG mill and inform strategies for enhanced mill efficiency.

### *B. Secondary Objectives*

1. Evaluate the correlation between feeder speed, level sensor data, and additional operational variables to identify the most influential factors affecting particle size distribution (PSD) before entering the SAG mill.
2. Characterize the variability in particle size distribution (PSD) from available ore feed data to assess its impact on feed consistency and its implications for the SAG mill's operational stability.
3. Validate PSD predictions derived from operational variables and evaluate their performance using statistical metrics in the absence of laboratory measurements.

## E. Dissertation Format

This dissertation consists of two research manuscripts, which are included at the end of this document as appendices. The following is a summary of the objectives and relevant findings of each manuscript.

*Manuscript #1 – "Blending Characterization for Effective Management in Mining Operations"*

The first manuscript (Appendix A) presents a comprehensive literature review on ore blending strategies, stockpile management, and the integration of data-driven techniques in mining operations. It explores the underlying causes of material variability, the operational challenges associated with achieving consistent feed characteristics, and the growing role of machine learning in improving blending outcomes.

This review addresses several key questions fundamental to understanding and designing predictive systems, including the variables most commonly linked to blending efficiency, the sensor technologies used to monitor ore properties in real time, and the computational methods applied to forecast process behavior. By synthesizing findings across both traditional and modern approaches, the paper provides practical insights for researchers and practitioners, supporting the development of more consistent, efficient, and adaptive blending systems in mineral processing.

*Manuscript #2 – "Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed"*

The second manuscript (Appendix B) focuses on the methodology, development, and validation of a data-driven framework designed to predict particle size distribution (PSD) in real time and support decision-making in SAG mill feed management. Building on insights gained from the literature review in Manuscript 1, this paper presents a detailed description of the modeling pipeline, including data preprocessing, clustering analysis, regression modeling, and sensitivity simulations.

The study was implemented as a web-based decision-support tool that enables operators to explore PSD predictions based on operational inputs and evaluate the feasibility of meeting specific target values. Model performance was validated using two years of operational data from a copper mine, achieving high predictive accuracy across multiple PSD targets. Additionally, a sensitivity analysis was conducted to quantify the influence of individual feeder adjustments on PSD, and a variability assessment compared PSD distributions under

historical versus model-guided scenarios, highlighting the potential of data-driven methods to improve feed consistency and support more stable grinding operations.

# Chapter 2

## Conclusions

This thesis assesses the role of data-driven technologies in improving particle size distribution (PSD) prediction and blending strategies within mineral processing circuits. Through the two manuscripts presented, the study addresses both the theoretical and practical aspects of feed variability management. The first manuscript provides a comprehensive review of blending practices, ore characterization methods, and machine learning applications, highlighting current limitations in predictive capabilities and the growing need for adaptive strategies to manage variability in mining operations. Building on these insights, the second manuscript develops and validates a real-time PSD prediction model based on historical operational data and machine learning techniques.

The results demonstrate that operational variables such as feeder speeds, stockpile levels, and throughput indicators can be effectively used to estimate PSD prior to milling. This data-driven case study reveals the strong predictive performance of Random Forest regression models across varying feed conditions. Additionally, sensitivity analyses offer valuable insights into how individual operational setpoints influence PSD outcomes.

By integrating prediction and sensitivity assessment into a user-accessible web application, this research illustrates the potential to shift from reactive to more proactive operational strategies in SAG mill circuits. While the approach has yet to be tested in real-time plant operations, it provides useful insights and practical tools that could support metallurgists, plant engineers, and decision-makers in improving grinding stability, energy efficiency, and sustainability in mineral processing.

**From Manuscript #1 – *"Blending Characterization for Effective Management in Mining Operations"***

This paper emphasizes that interest in data-driven blending strategies and predictive modeling has grown substantially in recent years, yet significant opportunities remain for advancing the field. The study provides a detailed overview of blending practices, ore characterization techniques, and the use of machine learning in mining. It addresses key questions concerning sensor technologies, blending effectiveness, and the computational approaches needed to model PSD behavior and reduce feed variability.

The findings underscore the potential of regression models and hybrid approaches to forecast PSD and support real-time operational decisions. By highlighting the relevance of sensor-based data and predictive analytics, this review contributes foundational insights for the design and implementation of more stable and responsive comminution systems.

**From Manuscript #2 – *"Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed"***

This paper presents a data-driven case study for modeling and managing particle size variability at the SAG mill feed. Using two years of high-frequency operational data from a copper mining operation, the study combines unsupervised clustering and Random Forest regression to generate accurate PSD predictions and evaluate the impact of feed adjustments.

The study highlights the value of sensitivity and variability reduction analysis in understanding the influence of operational variables on particle size distribution (PSD). Results demonstrate that feeder speeds have a measurable and often asymmetric impact on coarse PSD targets, with certain feeders exerting greater control potential than others. By quantifying these effects through univariate sensitivity metrics, the analysis provides actionable insights into which variables most effectively influence PSD outcomes. Furthermore, comparisons between historical and model-guided scenarios show that targeted adjustments can lead to significant reductions in output variability, particularly for coarser size fractions, supporting more stable and efficient grinding performance. These findings confirm the prac-

tical utility of predictive modeling and sensitivity analysis in improving process stability, enhancing energy efficiency, and advancing more sustainable practices in mineral processing.

# Chapter 3

## Future Works

A primary next step is to validate the PSD prediction models using live, real-time data from the site's operational infrastructure, such as data historians or industrial platforms. This real-time testing will assess model robustness under actual plant conditions, where noise and dynamics pose challenges absent in historical datasets, and will enable the transition from offline analysis to an operational decision-support tool.

Another important avenue is to expand the modeling framework by exploring deep learning approaches. More advanced models may capture complex nonlinear relationships and temporal dependencies more effectively, potentially improving predictive accuracy and adaptability in variable ore conditions.

Although this thesis focused on univariate sensitivity and variability analyses, a formal non-linear controllability assessment remains to be conducted. Future work should investigate multivariate control feasibility using data-driven methods adapted from control theory, such as reachability and observability analyses for non linear models. This will provide critical insights into the achievable PSD profiles given operational constraints and guide the design of more sophisticated control strategies.

Finally, the insights generated by this tool can directly inform and improve SAG mill feed control strategies. Integrating model predictions into plant control logic or operator advisories can support a shift from reactive to predictive feed management, enhancing process stability, energy efficiency, and overall grinding performance.

## References

[1] Pedro Lopez, Ignacio Reyes, Nathalie Risso, Moe Momayez, and Jinhong Zhang. Machine learning algorithms for semi-autogenous grinding mill operational regions' identification. *Minerals*, 13(11):1360, 2023.

[2] René Gómez, Krzysztof Skrzypkowski, Manuel Moncada, Raúl Castro, and Rodrigo Lazo. Segregation modeling in stockpile using discrete element method. *Applied Sciences*, 12 (23):12449, 2022.

[3] Z Ye, M Yahyaei, M Hilden, and MS Powell. Novel size segregation indices for multi-sized particle stockpiles. *Minerals Engineering*, 201:108165, 2023.

[4] GR Ballantyne and MS Powell. Benchmarking comminution energy consumption for the processing of copper and gold ores. *Minerals Engineering*, 65:109–114, 2014.

[5] Guohu Luo. Online detection system for ore particle size distribution based on deep learning. *Journal of Computer, Signal, and System Research*, 2(3):1–16, 2025.

[6] Paulina Quintanilla, Francisco Fernández, Cristobal Mancilla, Matías Rojas, Mauricio Estrada, and Daniel Navia. Digital twin with automatic disturbance detection for real-time optimization of a semi-autogenous grinding (sag) mill. *arXiv preprint arXiv:2407.06216*, 2024.

[7] Zahra Ghasemi, Mehdi Neshat, Chris Aldrich, John Karageorgos, Max Zanin, Frank Neumann, and Lei Chen. Enhanced genetic programming models with multiple equations for accurate semi-autogenous grinding mill throughput prediction. In *2024 IEEE Congress on Evolutionary Computation (CEC)*, pages 1–9. IEEE, 2024.

# Appendix A

## Blending Characterization for Effective Management in Mining Operations

Submitted to the *Mining, Metallurgy & Exploration* journal

**Matias Saavedra, Nathalie Risso, Moe Momayez, Ricardo Nunes, Victor Tenorio**

Department of Mining, Geological & Geophysical Engineering,
University of Arizona, 1235 James E. Rogers Way, Tucson, AZ 85719
United States

E-mail: matiassaavedra@arizona.edu
(520) 351 9225

*Article*

# Blending Characterization for Effective Management in Mining Operations

**Matias Saavedra [1], Nathalie Risso [1,*], Moe Momayez [2], Ricardo Nunes [2], Victor Tenorio [2], Jinhong Zhang [3]**

[1,2,3] Department of Mining and Geological Engineering, University of Arizona, 1235 E. James E. Rogers Way, Tucson, 85721, Arizona, USA
* Correspondence: nrisso@arizona.edu

**Abstract:** Ore blending plays a critical role in ensuring feed consistency and optimizing downstream processes in the mining industry. Despite its importance, achieving effective blending remains challenging due to ore variability and operational limitations. This review synthesizes current blending methodologies, covering both conventional techniques and modern, data-driven approaches. Special attention is given to the integration of Data Science and Machine Learning (ML) into blending strategies, demonstrating their effectiveness in predicting key process variables and supporting real-time decision-making. Key challenges, including data characterization, are discussed alongside enabling technologies such as sensor integration and automation platforms. By consolidating recent literature and highlighting emerging trends, this work outlines future directions for advancing intelligent blending systems and underscores the importance of standardized, high-quality data acquisition in mineral processing.

**Keywords:** Blending operations; Stockpile management; Machine Learning; Ore characterization.

## 1. Introduction

Blending quality is a critical factor in mining operations, directly influencing the efficiency and effectiveness of downstream processes. It encompasses various characteristics, including ore grade, Particle Size Distribution (PSD), moisture content, mineralogical composition, and mechanical properties such as hardness and abrasiveness. Variability in these attributes can disrupt process stability, leading to inefficiencies, increased energy consumption, equipment wear, and suboptimal product quality. For instance, ore grade determines the concentration of valuable minerals and impacts recovery, while PSD affects comminution efficiency and throughput [1]. Similarly, moisture content influences material handling and processing, potentially causing blockages in feeders or conveyors or affecting residency times [2]. Mineralogical composition dictates the selection of processing methods and reagents, while variations in ore competence or hardness can increase grinding energy demands and wear rates [3]. Managing these variations effectively is crucial for stabilizing operations, improving productivity, and ensuring sustainable resource extraction.

Beyond operational improvements, blending also contributes to sustainability efforts in mining. More consistent feed reduces energy consumption, equipment wear, and environmental impacts. The ability to predict and control ore variability also can improve resource planning, reducing risks associated with extraction and enhancing long-term viability. As mining operations strive to meet global sustainability goals, effective blending strategies are becoming essential for both operational and environmental objectives.

Received:
Revised:
Accepted:
Published:

**Citation:** Lastname, F.; Lastname, F.; Lastname, F. Blending Characterization for Effective Management in Mining Operations. *Journal Not Specified* **2025**, *1*, 0. https://doi.org/

Recent technological advancements have transformed blending from a reactive process into a proactive, data-driven approach. Real-time monitoring and sensor-based systems provide continuous feedback on ore characteristics, allowing dynamic adjustments to blending ratios. Machine learning (ML) algorithms and data-driven methodologies further enhance blending operations by identifying patterns and predicting future behavior. By integrating these technologies, mining operations can optimize blending decisions, reduce variability, and improve overall process performance [4].

This work reviews and analyzes the most widely adopted blending techniques in mining, ranging from traditional approaches adjustments based on operator expertise, to advanced data-driven and AI-based methodologies. The paper assesses the challenges, effectiveness, and limitations of blending approaches and explores opportunities to refine strategies through emerging technologies. By focusing on blending strategies and their impact on mining efficiency, this study provides a structured overview of current practices and future directions.

## 2. Background

### *2.1. Mining Operations and Blending*

Mining involves various stages to transform raw ore into valuable minerals, starting with **exploration** to identify mineral deposits, followed by **extraction** methods. Accurate identification and evaluation of mineral deposits are critical for successful project planning [5]. After extraction, ores move to the **blending** phase, where ores are mixed to create a more uniform feed for the processing plant. This step ensures that downstream processes, such as comminution, flotation, and refining, operate efficiently, maximizing recovery and reducing energy consumption [6]. Several blending methods are employed in mining to reduce feed variability and improve downstream process stability. These techniques involve the deliberate mixing of ores with differing properties to meet specific processing targets. Common blending methods include:

- **Wheeled Blending**: This method occurs typically at the stockyard or near the crusher pad. Ore is dumped in layers from different sources, and wheel loaders or bulldozers are used to mechanically mix the material before it is loaded onto conveyors or into crushers. It is often used when stockpiles receive material from multiple mining areas with varying characteristics.
- **In-pit Blending**: Ores from different zones within the pit are selectively mined and combined at the extraction face or directly in haul trucks. This allows early homogenization and reduces variability before the material leaves the pit.
- **Rotary Blending**: Conducted near the processing plant, this method uses rotating drums or dedicated mechanical systems to thoroughly mix ore batches in a controlled environment. It offers high precision and repeatability.
- **Conveyor Belt Blending**: Ore from different sources is fed at controlled rates onto a common conveyor system, allowing continuous blending en route to the plant. Feed rate control is key to achieving target blend ratios.

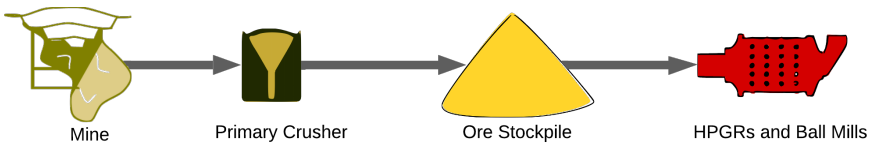


**Figure 1.** Mining Circuit Diagram

### 2.2. Operational Variables and Their Impact on Blending

While the importance of ore blending is well established, achieving consistency in practice requires control over key operational variables, including ore grade, moisture content, PSD, and flow rate. Equipment performance and stockpile reclaim strategies further influence blending outcomes, with poor stockpile management often leading to segregation and feed variability. Moisture, in particular, affects material flow and blending efficiency [7]. To mitigate these challenges, many operations have implemented automated blending systems with sensor-based feedback and computer-controlled platforms, allowing dynamic adjustment of blend ratios [8]. Consistent blending has been shown to enhance recovery, reduce energy consumption, and lower operating costs [9–11].

### 2.3. Ore Characterization and Mineralogical Analysis

Ore characterization through geo-characterization plays a key role in managing variability and provide meaningful information to support resource estimation and mine planning [12]. Techniques like Scanning Electron Microscopy (SEM) and X-ray Diffraction (XRD) are crucial for understanding ore composition, which can help predict processing outcomes and improve recovery [13]. Comprehensive ore characterization is vital for identifying valuable minerals and undesirable elements, which leads to better processing outcomes and informed decision-making [13]. Additionally, PSD analysis is a key factor in improving grinding circuits and flotation efficiency, thus improving recovery [14]. Combining PSD with mineralogical data enhances crushing and flotation efficiency, which directly contributes to higher recovery [15].

### 2.4. Advanced and Real-Time Techniques

Real-time techniques, including sensor-based systems, are instrumental in improving adaptive operations in mining, as they allow for continuous monitoring and adjustment of critical variables such as PSD, ore composition, and moisture content [15]. Technologies such as Near-Infrared (NIR) spectroscopy, X-ray fluorescence (XRF), laser scanning, and online image-based particle size analyzers are increasingly being deployed to provide rapid, in-situ measurements. These systems feed data into control architectures and digital platforms, enabling real-time feedback loops that inform blending decisions and process set points. For instance, real-time PSD data can be used to adjust feeder rates or modify mill speeds, while continuous grade measurements support dynamic stockpile management. The integration of these sensor technologies with machine learning algorithms and process control systems not only improves the speed and accuracy of decision-making but also reduces variability, enhances ore quality consistency, and optimizes resource use.

SCADA (Supervisory Control and Data Acquisition) platforms further enable centralized oversight of sensor networks and equipment. In blending applications, data from feeders, conveyors, and analyzers is streamed into SCADA, where it is aggregated, visualized, and used to trigger automated control actions. These systems support real-time feedback loops by allowing operators—or connected control algorithms—to adjust setpoints dynamically based on changing ore conditions. When integrated with digital twins and ML-based forecasting models, SCADA systems form the backbone of more adaptive and autonomous blending strategies.

## 3. Literature Review

This literature review seeks to assess existing research on the application of data driven, blending strategies, and ore characterization in mining operations. Our evaluation focuses on key aspects, data quality assessment, including methods and techniques for blending characterization, strategies for mitigating segregation in stockpiles, data-driven approaches

for ore characterization, and the impact of these processes on operational efficiency and sustainability.

The review adheres to the guidelines outlined in the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) checklist [16]. The PRISMA methodology considers

1. Defining the potential research questions from the problem domain and retrieve relevant studies from multiple databases.
2. Establishing the inclusion and exclusion criteria to filter the most relevant studies.
3. Extracting and compile key data from the selected studies to provide evidence-based answers to the research questions.

### *3.1. Research Questions*

We define a set of research questions aimed at guiding our exploration of ore particle size distribution, stockpile management, instrumentation, and predictive methods in mining operations.

- **Research Question 1:** What variables are used to define blending criteria?
- **Research Question 2:** What sensors and technologies help monitor critical variables for blending?
- **Research Question 3:** What performance metrics and evaluation methods are being used to assess blending effectiveness?
- **Research Question 4:** In which areas of the blending process is machine learning applied, and which models are commonly used for each task?

### *3.2. Search Strategy*

The literature reviewed was sourced from major academic databases, including ScienceDirect, Elsevier, Springer, and IEEE Xplore, with a focus on high-impact journals in mining, geotechnical engineering, and data-driven mineral processing. Key journals include *Minerals Engineering*, *International Journal of Mineral Processing*, *Mining Technology*, *Minerals (MDPI)*, *Rock Mechanics and Rock Engineering* and *IEEE*, among others selected for their relevance to blending and process optimization.

The search string was developed by combining keywords based on the pre-defined research questions. The initial query used was: *"blending" AND "data" AND "mining" AND ("analytics" OR "modeling" OR "machine learning" OR "artificial intelligence" OR "AI" OR "data analysis" OR "statistical analysis")*. After selecting relevant papers, we refined the focus to explore a broader area. For this purpose, we used the following search string: *"blending" AND "particle size" AND "distribution" AND "mining"*. To complete the literature review and gain broader insights into the influence of specific parameters, we then used this set of keywords: *"blending" AND "mineral" AND "characterization" AND "mining"*. This search resulted in a total around almos 40 papers from the journals mentioned.

As shown in Figure 2, the distribution of publications shows a peak in 2021, indicating a notable increase in research activity on blending operations characterization within the mining industry during that year enabled by recent advancement in data-driven strategies.

### *3.3. Inclusion and Exclusion Criteria*

To refine the selection, the authors defined clear inclusion and exclusion criteria to narrow the focus of the literature review. Inclusion criteria are designed to focus the review on papers directly related to key topics of interest, such as blending operations, ore characterization, particle size distribution, and their impact on process efficiency in mining. The following criteria are used for paper selection:

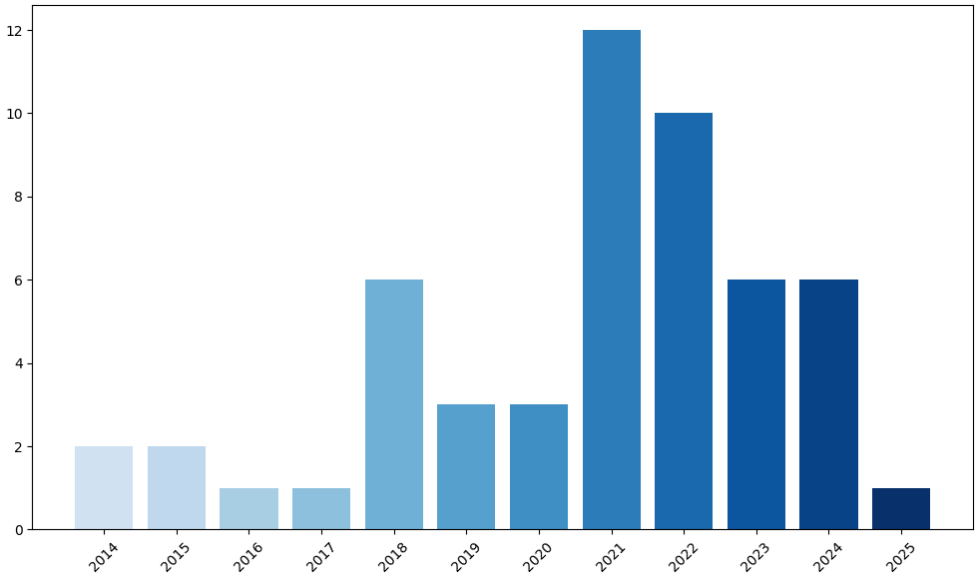


**Figure 2.** Distribution of papers by year.

1. Blending Operations, Particle Size Distribution, and Ore Characteristics in Mining: Papers that discuss blending operations, including how PSDand ore characteristics influence blending decisions and downstream processes in mining operations.
2. Blending Efficiency and Operational Impact: Papers that focus on evaluating and improving blending efficiency in mining operations, particularly by analyzing how blending practices influence ore consistency, downstream processing, and overall production performance.
3. Machine Learning and Predictive Models for Blending: Papers that develop or apply machine learning, data-driven techniques, or predictive models specifically to support blending decisions, ore characterization, and real-time process control.
4. Time Span: Papers published within the range from 2014 to 2025, ensuring that the review captures the most recent and relevant research on blending operations and associated techniques in the field.

For the exclusion criteria we focused on the following aspects:

1. Incompleteness of Inclusion Criteria: Any paper that did not meet the specified inclusion criteria, was discarded.
2. Focus on Metallic Ores in Open-Pit Mines: Studies not involving metallic ores or not clearly focused on open-pit operations were excluded.

## 4. Results and Discussion

In this section, we present the results of our comprehensive review, which aims to address the research questions outlined in Section 3.1, followed by a discussion on the findings. The ordering of studies in Tables (1) and (2) reflects a thematic grouping consistent with section (4). Studies are grouped based on their focus on ore variability, real-time modeling strategies, or sustainability considerations throughout the mining process, rather than by publication date.

### *4.1. Blending and Ore Variability*

As shown in Tables (1) and (2), blending plays a critical role in stabilizing ore composition and improving downstream processing performance. Studies in Table (1) (rows 1–7) highlight the need to minimize variability in ore type, particle size distribution (PSD), and

flow rates, all of which impact blend uniformity. For example, [17] and [18] examine how fluctuations in feed composition affect flotation efficiency and comminution performance, reinforcing the importance of consistent blending to maintain process stability.

### *4.2. Real-Time Monitoring and Sensor Technologies*

Effective blending relies on continuous monitoring of ore properties. Commonly used technologies include image-based PSD analyzers, microwave or NIR moisture sensors, and online grade analyzers such as PGNAA and LIBS, often paired with belt scales for mass-based tracking. These sensors provide the basis for real-time quality calculations, where values such as weighted ore grade or PSD averages are computed continuously based on sensor inputs and flow rates. These calculations enable timely adjustments to feeder speeds or blending ratios.

The author in [19] demonstrates this through a sensor-driven blending system that updates stockpile quality in real time. Similarly, [20,21] apply heuristic and differential evolution algorithms to maintain optimal blends under changing conditions, and [22] uses linear programming to stabilize grade delivery.

### *4.3. Computational and Modeling Approaches*

To address the inherent complexity of blending and stockpile management, several studies (Table (1), rows 8–15) adopt simulation-based and data-driven modeling approaches, targeting different stages of the mining process. [23,24] employ 3D cellular automata to simulate stockpile formation and internal segregation, supporting real-time adjustments during material stacking. These models are particularly useful at the interface between truck dumping and conveyor distribution systems.

The research conducted on [25] implements digital twins to enable continuous material tracking across crushers, conveyors, and stockpiles. This allows scenario analysis and proactive decision-making before material reaches the mill. in a similar line, [26] integrates ultrasonic sensors with machine learning to control mixing at the reclaim feeder stage, where blended material is retrieved from the stockpile for SAG mill feed.

Other contributions, such as [27,28], focus on feed prediction and reconciliation using geostatistics and sensor fusion. These techniques are commonly implemented during ore dispatch and stockpile reclaim, helping to align planned and actual feed conditions. Altogether, these computational frameworks enhance blending adaptability and precision, especially when integrated with real-time feedback from SCADA systems and online analyzers.

### *4.4. Operational Constraints and Sustainability Challenges*

Predictive modeling is increasingly used in blending to forecast key outcomes such as feed quality, recovery, and PSD stability. However, its application in real operations is often limited by ore variability, fragmented data, and integration challenges. Studies listed in Table 2 (rows 16–19) examine these constraints from both operational and environmental perspectives.

The author in [29] proposes a nonlinear model to manage short-term feed variability, while [30] highlights how ore fragmentation influences blending effectiveness, reinforcing the need to coordinate upstream and downstream decisions. [31] points to the long-term sustainability risks of poor material handling, including particle size segregation in waste piles.

Automation technologies can help address some of these challenges. Programmable logic controllers (PLCs) manage local control actions, while supervisory control and data acquisition (SCADA) systems centralize monitoring. Real-time sensors, such as level and flow transmitters, provide continuous feedback, and vision systems like PSD cameras

enable automated quality checks. [32,33] show how these technologies improve blending consistency, equipment use, and safety.

Predictive control, where model outputs are used to adjust operations in real time, offers further potential. Yet, its adoption remains limited, primarily due to system integration complexity and the need for transparent, operator-friendly interfaces.

**Table 1.** Overview of Blending Practices and Associated Challenges. Part 1

| No. | Author | Year | Objective/Goal | Methodology | Instrumentation | Results | Conclusion |
|---|---|---|---|---|---|---|---|
| 1 | Ozlem Bicak [17] | 2019 | To study ore variability in sulphide ores and its impact on flotation. | Oxidation Index using EDTA extraction. | ToF-SIMS, XPS | High Oxidation Index correlates with copper-enriched ore contamination and poor flotation recovery. | Oxidation Index is a practical, cost-effective tool for geometallurgical mapping and flotation prediction. |
| 2 | Yue Xie, Frank Neumann, Aneta Neumann [21] | 2021 | To optimize stockpile blending under uncertainty using chance-constrained programming. | Nonlinear optimization model with chance constraints, Differential Evolution (DE) algorithm. | No specific instrumentation mentioned. | DE algorithm outperforms deterministic models, with chance constraints affecting success rates. | Chance-constrained optimization effectively manages uncertainties, with copper grade constraints significantly influencing results. |
| 3 | Diego Marques, João Felipe C Costa [34] | 2013 | To simulate ore grade variability in blending and homogenization for improved processing efficiency. | Geostatistical simulations using the Turning Bands Method (TBA), 3D grade block model. | No specific instrumentation mentioned. | Homogenization reduces silica ($SiO_2$) variability, with effectiveness tied to pile size and layer count. | Optimal homogenization parameters are selected based on pile mass and layer count for better processing results. |
| 4 | Shi Zhao, Tien-Fu Lu, Ben Koch, Alan Hurdsman [35] | 2015 | To develop a real-time method for stockpile quality calculation. | 3D voxel model, laser scanning, BWR model. | Laser scanner, 3DOF scanner, field trials for UKF localization. | 3D voxel model improves stockpile quality and geometry predictions. | Limitations: scanning accuracy, high computation time, and BWR reclaiming pattern. Real-time chemical data is often unavailable. |
| 5 | Yue Xie [21] | 2021 | To improve large-scale stockpile blending and metal grades using the DE algorithm. | Continuous improvement model with DE algorithm and repair operators. | DE algorithm, repair operators. | DE algorithm improves metal volume over traditional methods. | DE algorithm enhances stockpile blending and mine scheduling efficiency. |
| 6 | Tatenda Huggins [18] | 2018 | To study ore blending effects on comminution and product quality at LKAB. | Investigates blending effects on grinding and product quality, focusing on high tonnage operations. | Laboratory-scale milling, energy consumption, particle size distribution. | Binary blends more efficient than tertiary blends. | Ore blending improves energy use, reduces maintenance, and enhances product quality. |
| 7 | Mohamed Farghaly [22] | 2021 | To optimize blending and production processes using linear programming. | Linear programming model to optimize costs and production readiness. | Linear programming model, LINDO TM software. | Optimal model reduced costs to 194.183 million, highlighting key decisions for blending, batch size, and transportation. | Model improves profitability, verified with Cu and Co production data. |
| 8 | Z. Ye, M. Hilden, M. Yahyaei [24] | 2022 | To develop a 3D cellular automaton model for simulating ore pile formation and segregation. | 3D CA model for ore pile formation, size segregation. | No specific instrumentation. | Model simulates ore pile formation and segregation; validated with laboratory-scale data. | Limitations include inability to represent conical piles and neglect of compaction effects. |
| 9 | Z. Ye, M.M. Hilden, M. Yahyaei [23] | 2023 | To develop a fast 3D cellular automata-based stockpile model for process control. | 3D CA model for real-time stockpile monitoring and size segregation. | Ultrasonic height sensors, machine vision-based particle-size analysis. | Model simulates stockpile behavior accurately, validated with industrial data. | Enhances process control, but assumptions about free-flowing ore and compaction limit its effectiveness. |
| 10 | Martin Servin, Folke Vesterlund, Erik Wallin [25] | 2021 | To explore digital twins with particle simulation for material tracking. | Digital twins, distributed particle simulation for tracking material in mine-to-mill processes. | Sensors, simulation models, AGX Dynamics, MWD data, vehicle telematics. | Particle-based digital twin improves material tracking and process efficiency. | Limitations in supporting size segregation and capturing stockpile surface shape. |
| 11 | Alexander Bowler, Serafim Bakalis, Nicholas Watson [26] | 2020 | To use ultrasonic sensors and machine learning for optimizing mixing processes. | Ultrasonic sensors and machine learning for non-invasive monitoring of mixing. | Ultrasonic sensors, machine learning, MATLAB R2019a. | High accuracy in classifying mixing states and predicting completion times. | Challenges in obtaining labeled data and achieving optimal performance across algorithms. |
| 12 | J Wu [36] | 2022 | To develop an open-pit automation ore blending method. | Constraint model for ore-matching, improving blending and reducing grade deviation. | Genetic algorithm, geostatistics for grade prediction, automated blending decisions. | Model improves ore quality and production rates. | Enhances efficiency and reduces costs in open-pit mining. |
| 13 | Jörg Benndorf [27] | 2016 | To improve mine production control with sensor data and geostatistical modeling. | Framework for reducing uncertainties and optimizing real-time decisions. | Sensor-based characterization, geostatistical modeling, Kalman Filter. | 15-40% improvement in production control and reduced uncertainty. | Reduces losses from discrepancies in production targets. |
| 14 | William Reid [37] | 2021 | To study uncertainty in advanced ore mining using Maptek's Evolution system. | Evolutionary computation and neural networks for quantifying uncertainty in extraction. | Evolutionary computation, geological uncertainty, staging approaches. | Demonstrates how staging methods impact NPV and profit. | Improved profitability through geological uncertainty modeling. |
| 15 | T Wambeke, J Benndorf [28] | 2018 | To study the influence of blending ratio and sensor precision on grade control models. | Simulation-based geostatistical approach for grade control model reconciliation. | Sensors for measuring blended material streams, varying measurement errors. | RMSE reductions of 20-60% in grade control models. | Highlights the importance of sensor accuracy and blending ratios. |

**Table 2.** Overview of Blending Practices and Associated Challenges. Part 2

| No. | Author | Year | Objective/Goal | Methodology | Instrumentation | Results | Conclusion |
|---|---|---|---|---|---|---|---|
| 16 | Pedro Henrique Alves Campos [29] | 2024 | To optimize mine scheduling and metallurgical recovery with nonlinear blending models and simulated annealing. | Nonlinear models, simulated annealing. | Simulated annealing, blending models. | Improves recovery and efficiency. | Penalty factor lacks physical meaning, and linear models reduce effectiveness. |
| 17 | Peiyong Qiu, Thomas Pabst [31] | 2023 | To analyze particle size segregation in a waste rock pile. | Image analysis for segregation. | Drone, 87 photos, waste rock pile. | Segregation increases from top to bottom, with lateral heterogeneity. | Impacts hydraulic conductivity and geotechnical properties. |
| 18 | Ayyub Nikkhah [30] | 2022 | To optimize ore fragmentation from blasting at Sarcheshmeh copper mine. | Blasting parameter optimization. | Statistical analysis, fragmentation efficiency. | Tailored blasting improves performance and reduces costs. | Enhances mining efficiency, further research needed. |
| 19 | Loreto Codoceo-Contreras [33] | 2024 | To explore automation impacts in mining through NLP review of 94 documents. | NLP review, thematic analysis. | 94 documents, TF-IDF, anomaly detection. | Identifies themes like workload and communication. | Recommends research on community impacts and cybersecurity. |

### *4.5. Discussion*

From Tables 4 and 5, we identify key machine learning trends and methodologies aligned with the research questions outlined in Section 3.1. These studies reflect the growing use of ML models in ore blending and mineral processing to support real-time prediction and control of variables such as throughput, grinding efficiency, and particle size distribution. Algorithms like regression models, neural networks, gradient boosting, and hybrid optimization methods are applied to enhance blending performance under dynamic conditions.

Machine learning and artificial intelligence (AI) continue to expand their role in modern mining operations [38], offering tools for predictive modeling, process optimization, and sustainability improvement. These applications highlight the shift from reactive to data-driven decision-making in mining.

Several studies adopt ensemble and deep learning techniques to model nonlinear process behavior and improve feed quality. The use of hybrid methods, combining models like CatBoost with evolutionary optimization, or deep learning architectures such as LSTM networks, has shown potential for handling time-series data and capturing long-term dependencies. These approaches are particularly useful in grinding circuits and dynamic blending scenarios.

**Table 3.** Machine Learning Models for Blending and PSD Prediction.

| Model / Method (with citation) | Input Variables | Predicted Output | Performance Metrics |
|---|---|---|---|
| Artificial Neural Network (ANN, 3-10-1) [39] | Rod load (%), Feed load (%), Grinding time (min) | PSD (percent passing per sieve) | R = 0.999 (train), R = 0.987 (val), RMSE = 0.165–0.965 |
| ANN + Genetic Algorithm [40] | Feed size, Ore hardness, Liner wear | Energy consumption | RMSE reduced vs. baseline; improved accuracy |
| Artificial Bee Colony (ABC) Optimization [36] | Ore grades, costs, recovery constraints | Optimized ore blending plan | Recovery: W +8.95%, Mo +12.2%, Bi +8.33%; Grade fluctuation <10% |
| Goal Programming Optimization [41] | Rock powder sampling, Grade priority constraints | Optimized ore blending schedule | Grade deviation minimized; rapid linear optimization |
| Stochastic Gradient Boosting (SGB) [42] | MIR/NIR spectra, soil C, pH, RGB, extractants | PSD (sand, silt, clay %) | $R^2$ = 0.87–0.95 (MIR); $R^2$ = 0.66–0.79 (NIR) |
| Deep LSTM (DRNN) [43] | Ore feed, Water flow (sumps), Control loops | 6 KPIs (Throughput, Solids %, Size fractions, Recirculation) | Accuracy = 99%; outperforming wavelet models |
| CatBoost + Differential Evolution [44] | SAG mill weight, power draw, feed size, water rate | SAG mill throughput (t/h) | Highest prediction accuracy (CatBoost); best robustness (DE) |
| ANN, Random Forest, XGBoost, GBM [45] | SAG feed P80, granulometry (<30 mm, >100 mm), rotational speed, liner age, solids %, stockpile level | SAG mill production (TpH), energy consumption | ANN: $R^2$ = 0.885, RF: $R^2$ = 0.755, XGBoost: $R^2$ = 0.772; Production increase 4.42%, Energy decrease 7.62% |

The models reviewed in Table 3 represent a wide range of approaches to solving key operational problems in mineral processing, particularly in predicting particle size distribution (PSD), optimizing ore blending, and forecasting throughput. Across these

studies, performance indicators such as high $R^2$ values and reduced RMSE confirm the capability of machine learning techniques to extract meaningful patterns from complex operational data.

In PSD prediction, models like artificial neural networks (ANN) and long short-term memory (LSTM) networks have achieved high precision under controlled conditions [39,43]. These models effectively capture non-linear dependencies between variables such as grinding time, feed characteristics, and size output. However, the datasets used in these studies are often well-structured, with consistent sampling and limited noise—conditions not always mirrored in plant environments. In industrial settings, factors like ore variability, sensor drift, and asynchronous measurements introduce noise that can compromise model robustness.

For blending optimization, algorithms such as Artificial Bee Colony and Goal Programming offer structured frameworks to support decision-making under constraints [36,41]. These approaches focus not just on prediction but on selecting input combinations that meet multiple objectives—such as grade control, recovery, and cost minimization.

Among more general-purpose learning methods, Random Forest strikes a practical balance between predictive accuracy and usability in real environments. Its tolerance for missing or noisy data, combined with minimal preprocessing requirements, makes it particularly attractive for deployment in supervisory applications. In [45], Random Forest and related models were successfully applied to SAG mill data to improve throughput prediction and energy efficiency, demonstrating the model's adaptability under variable conditions.

Despite these promising results, the deployment of machine learning models at scale is often limited by inconsistent data practices. Studies such as [39,44,45] differ widely in their use of variables, measurement units, preprocessing steps, and data quality control. These discrepancies hinder not only model transfer between operations but also comparative benchmarking and reproducibility. Even models with strong local performance may fail to generalize if retrained on incompatible datasets. Addressing this issue will require industry-wide adoption of standardized data acquisition protocols, consistent variable naming conventions, and clear preprocessing guidelines. These steps are essential to enable scalable, trustworthy, and transferable predictive tools for blending control.

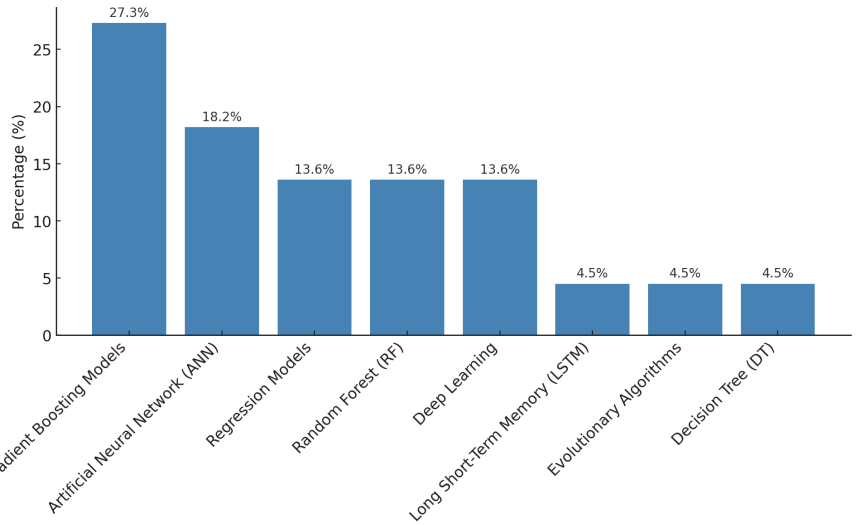


**Figure 3.** Frequency of Machine Learning Models in Blending

Among the most frequently applied machine learning models in blending and quality control, Artificial Neural Networks (ANN), Random Forest (RF), and Gradient Boosting

Machines (GBM) are particularly prominent. These models are widely adopted for predicting PSD, throughput, and recovery, offering a combination of flexibility and strong predictive performance. ANN and RF, in particular, have shown consistent success across multiple case studies, demonstrating their reliability in both structured and more complex operational contexts. Moreover, deep learning techniques such as Long Short-Term Memory (LSTM) networks are gaining traction due to their capacity to model temporal dependencies and nonlinear dynamics in ore processing. As machine learning continues to evolve, hybrid approaches combining deep learning and ensemble methods are expected to further enhance accuracy and robustness, contributing to more efficient and adaptive blending optimization strategies across mining operations.

**Table 4.** Machine Learning Applied to Blending. Part 1.

| N | Author | Objectives | Methodology | Instrumentation | Data Used | Key Variables | Results |
|---|---|---|---|---|---|---|---|
| 1 | Zongnan Li [42] | Establish PSD model for mine tailings and validate with samples. | Mathematical PSD model with A, B, K coefficients; validated with twelve tailing materials. | PSD determined using Laser PSA | Twelve tailing samples from metal mines in China. | Model coefficients (A, B, K) for PSD. | $R^2 > 0.99$ for all twelve PSD lognormal curves. |
| 2 | Srinivas Soumitri Miriyala [43] | Develops a multi-objective optimization algorithm for LSTM networks to improve generalization | Focuses on nonlinear system identification in grinding circuits, comparing optimal LSTMs with traditional tools | Deep recurrent neural networks for system identification, highlighting data model accuracy needs | Lack of sensors in circuits, analyzing overflow slurry properties for KPIs | Data from PRBS signals; LSTM networks achieved 99% accuracy, tested on unseen signals | Emphasizes throughput as a productivity indicator and evaluates models using AIC and RMSE |
| 3 | Brian Lindner [46] | Develops a fault diagnosis approach using process topology and feature extraction for fault detection | Applies methodology to industrial concentrator data to create a connectivity map | Combines process topology and feature extraction for effective diagnosis | Uses historical data to extract topology, generating a connectivity map for analysis | Key variables include Mill2 and Mill3; identifies root nodes influencing cyclone feed | AUCs improved from 0.85 to nearly 1; methodology emphasizes accurate fault detection and topology integration |
| 4 | Nartey Faustin [47] | Create ML models to predict mine production and enhance efficiency. | ANN, RF, GBR, and DT models; MLR as baseline; regression analysis for feature selection. | Not addressed | 126 historical datasets from Pit W of Mine X. | Daily trucks, excavator hours, and utilization. | ANN model best ($R^2$ = 0.8003); MLR worst ($R^2$ = 0.6044). |
| 5 | Bingyu Liu [48] | Establish a quality model for ore blending to enhance product quality. | Intelligent ore blending method using NSGA-II and ABC algorithms. | Not addressed | Historical blending and dressing data. | Ore grade, high-quality content. | New method improved profits. |
| 6 | F Nakhaei [49] | State-of-the-art 3D imaging in mineral processing, focusing on automation and efficiency | 3D imaging techniques (RhoVol, X-ray CT), high-speed scanning, integration for automation | RhoVol measures particle mass, XCT captures internal structures, improving mineral characterization | 3D imaging for particle size, shape, and composition analysis, comparison with sieving methods | Key variables: size, shape, porosity, crack density, and intergranular fractures, essential for characterization | Not addressed |
| 7 | Shi Zhao [35] | Develop a real-time 3D volumetric model and improve quality grade accuracy. | Triangular prism model, sub-prism partitioning, adaptive reclaiming, and blending. | LMS200 & O1D100 laser scanners | 2D laser range finder data for stockpile profiling. | Stockpile shape, quality grades, angle of repose, density. | $R^2$ for regression accuracy, SSE for model evaluation. |
| 8 | Ji Wu [36] | Improve ore blending to improve recovery and cost efficiency. | ABC algorithm for multi-objective optimization; software developed for blending. | Algorithm implemented in Matlab | Historical data from Shizhuyuan Mine, focusing on ore recovery. | recovery, mining costs, production capacity. | Recovery rate increased from 0.45% to 0.491%. |
| 9 | Dahee Jung [50] | Review ML applications in mining and analyze trends. | Systematic literature review of 109 papers; keyword search; selection criteria defined. | Not addressed | Various datasets, including open-source data from RapidEye. | Data sources, utilization types. | ML performance evaluated; deep learning average RMSE: 0.000608. |
| 10 | Qiang Shi [51] | Review ML applications in open-pit mining to improve predictability and feasibility | Statistical analysis of 100+ papers (2012–2023), categorized into four key areas | Sensors for real-time drilling, hyperspectral data analysis | Operational resources, equipment utilization, mine geology, production, environment | Block extraction, tonnage (tons), volume (cubic meters), destination assignment | Not addressed |
| 11 | Amit Kumar [47] | Explore AI applications in mineral processing, addressing challenges and sustainability | Trends and methodologies in ML/AI, agile and spiral approach, data investigation | SensAI systems, low-fidelity sensors, hyperspectral imaging, ultrasound sensors | Sensory data from mineral processing plants, exploratory data analysis | Latent factors, process parameters, model parameters, feature engineering | Statistical analysis, data clustering, exploratory data analysis |
| 12 | Larissa Statsenko [52] | Develop an info system for ore blending to stabilize output quality. | Imitation modeling and simulation for ore blending design; decision support system applied. | Not addressed | Datasets from four open-pits in Kazakhstan. | Stockpile dimensions, transport mass, Fe-grade. | Ore grades variance reduced 4 times; blending coefficient improved 1.5-2 times. |
| 13 | Elizabeth [53] | Evaluates infrared (IR) models for soil PSD using machine learning (SGB) to predict sand, silt, clay, and carbon content. The 4X gain method achieves $R^2$ values between 0.87–0.95, NIR and MIR spectra give $R^2$ between 0.40–0.94. | SGB applied to NIR and MIR spectra; $R^2$ values: 0.40–0.94 for NIR, 0.45–0.80 for MIR, 4X gain method: $R^2$ = 0.87–0.95. | Stochastic Gradient Boosting (SGB) | 1298 soil samples; $R^2$ for sand sieving: 0.66, sedimentation: 0.53, NIR: 0.69–0.92, MIR: 0.45–0.80, SDF and settling time: 0.84–0.92. | Sand, silt, clay, carbon content, SDF, settling time, NIR, MIR spectra. | 4X gain: $R^2$ = 0.87–0.95; NIR and MIR: $R^2$ = 0.40–0.94; Settling time and SDF: $R^2$ = 0.84–0.92. Machine learning enhances PSD prediction. |
| 14 | Y. S. Kim et al. [39] | Investigates ANNs for predicting PSD to reduce energy consumption. | Supervised ANN trained on experimental data. | Rod mill for grinding, PSD measurement of coal. | Experimental data with varying feed vol.%, rod load vol.%, and grinding time. | Feed volume%, rod load, grinding time. | ANN predicted PSD well; RMSE ranged from 0.165 to 0.965, showing its ability to predict PSD under varying conditions. |

**Table 5.** Machine Learning Applied to Blending. Part 2.

| N | Author | Objectives | Methodology | Instrumentation | Data Used | Key Variables | Results |
|---|---|---|---|---|---|---|---|
| 15 | Shubham Shrivastava [54] | Develops an AI method to predict PSD from dump images. | Mask R-CNN with ResNet50 and FPN trained on dump images. | Mask R-CNN with ResNet50 and FPN for image segmentation. | Images from iron ore dump, particles 5 mm to larger. | Particle size distribution, image resolution, segmentation accuracy. | Mask R-CNN: 0.936 accuracy, 0.829 loss, predicts 250-300 particles/image, 1.274 mm error; predicts wide range of particles. |
| 16 | Both, Christian [40] | Predict throughput at Tropicana Gold Mining using machine learning. | Comparison of neural network models and linear regression for throughput prediction. | Data from blasthole drilling, comminution circuit, ball mill power, product particle size. | Production data including hardness proportions, ball mill power, particle size. | Hardness proportions, ball mill power, product particle size. | Neural network reduces RMSE by 10.6%, improves accuracy with ball mill power and particle size; hardness proportions enhance prediction by 6.3%. |
| 17 | Ghasemi, Zahra [44] | Develop a hybrid framework to optimize SAG mill throughput using machine learning and evolutionary algorithms. | Evaluated 17 models with 36,743 records, applying feature selection and three evolutionary algorithms (Differential Evolution, GA, PSO). CatBoost was the most accurate, and Differential Evolution was the best optimization algorithm. | SAG mill operational data including throughput and parameters. | 36,743 records from SAG mill operations. | CatBoost, Differential Evolution, throughput, parameters. | CatBoost achieved $R^2$ = 0.82, with Differential Evolution providing robust throughput predictions. Key metrics: RMSE = 49.6, MAE = 36.3, EVS = 0.645, $R^2$ = 0.82. |
| 18 | Saldaña, Manuel [45] | Optimize SAG mill grinding in Chilean copper mining using machine learning (ANN, random forest, GBM) to analyze the impact of fragmentation, mill power, and liner age on throughput. | Compares multiple models (ANN, random forest, GBM, XgBoost, linear regression) for throughput prediction and mill parameter optimization. | Production data from Chilean copper mining, including mill power, fragmentation, and liner age. | Production data, mill power, fragmentation, liner age, energy consumption, throughput (TpH). | ANN, Random Forest, GBM, XgBoost, Linear Regression | ANN performed best with $R^2$ = 0.89. GBM and XgBoost achieved $R^2$ = 75.46% and 77.18%, while linear regression had $R^2$ = 0.55. ANN increased production by 4.41% and reduced energy consumption by 7.62%. Machine learning models outperformed linear regression for throughput prediction. |

## 5. Conclusion

This review highlights the critical role of blending in improving downstream performance in mineral processing operations such as milling and flotation. The effectiveness of blending is influenced by key variables including ore type, grade, particle size distribution (PSD), moisture content, and flow rates. A multidisciplinary combination of advanced modeling, simulation, sensor integration, and predictive analytics is essential to achieve more consistent and optimized blending outcomes.

Simulation models, integer programming, and geometallurgical approaches have demonstrated success in stabilizing ore grades and minimizing variability. Meanwhile, sensor technologies such as Near-Infrared (NIR) spectroscopy, hyperspectral imaging, and X-ray fluorescence (XRF) enable real-time monitoring of ore properties, providing actionable feedback for dynamic control of blending operations. Machine learning (ML) methods—particularly artificial neural networks (ANN), random forest (RF), and decision trees (DT)—have shown significant promise for predicting throughput, PSD, and blending performance using both historical and real-time data. These models reduce reliance on manual heuristics, enhance decision-making, and improve resource efficiency. Hybrid approaches that combine deep learning with traditional ML methods are expected to deliver even more accurate and robust predictions in future implementations.

However, as observed across multiple studies, the lack of data standardization remains a significant barrier to wider adoption. Inconsistencies in data collection frequency, variable

labeling, and measurement techniques hinder the comparability and transferability of predictive models. Establishing standardized data acquisition protocols is essential for ensuring model reliability, facilitating cross-site deployment, and enabling more reproducible results.

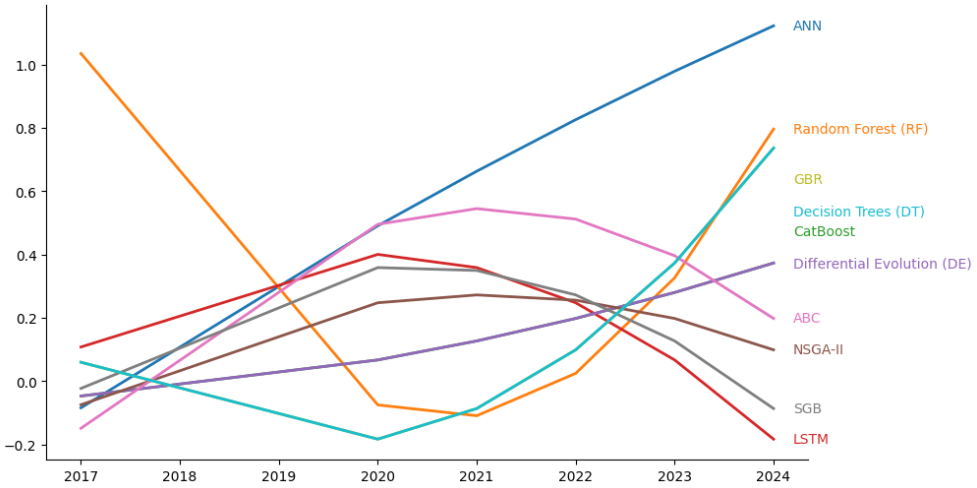


**Figure 4.** Trends in the Application of Machine Learning Methods for Blending Optimization

Figure 4 illustrates the increasing adoption of ML techniques in blending optimization, with ANN and RF leading in frequency and performance. These trends reflect a broader shift toward data-driven blending strategies that integrate real-time sensor feedback, automation platforms, and predictive control frameworks.

Looking ahead, future research should focus on refining ML models, integrating them with PI System infrastructure and real-time monitoring tools, and advancing fault detection capabilities. Prioritizing data standardization and transparency will be key to ensuring scalable and sustainable blending optimization across diverse mining contexts. By combining predictive technologies with robust automation, mining operations can achieve improved ore grade stability, reduced variability, and enhanced overall process efficiency.

## References


1. Prasojo, T.S.; Yulianto, A.; Hindarto, A.; Parinussa, B.; Arifien, A. Ore blending as mine scheduling strategy to accommodate resources conservation at pakal nickel mine, PT ANTAM (Persero) Tbk. *Procedia Earth and Planetary Science* **2013**, *6*, 24–29.
2. Rooplal, B. Designing an ore processing plant – Factors you need to consider. *Mineral Processing Insights* **2022**.
3. Giblett, A.; Putland, B. The basics of grinding circuit optimisation. In Proceedings of the Proceedings of the AusIMM Mill Operators' Conference, 2014.
4. Zhang, X.; Li, L.; Chen, Y. Data Mining Approaches for Optimizing Ore Blending Decisions. *Minerals* **2022**, *12*, 48. https://doi.org/10.3390/min12010048.
5. AO, D.W. Geology and Mining: An Introduction and Overview. *SEG Newsletter* **2018**, pp. 1–21.
6. Valery, W.; Duffy, K.; Jankovic, A. Mine to Mill Optimization. *SME Mineral Processing and Extractive Metallurgy Handbook* **2019**, pp. 335–343.
7. Nzulu, G.K.; Bakhit, B.; Högberg, H.; Hultman, L.; Magnuson, M. Elucidating Pathfinding Elements from the Kubi Gold Mine in Ghana. *Minerals* **2021**, *11*, 912.
8. Zhou, X.; Zhang, H.; Wu, Q. Optimal Blending of Polymetallic Ore Using an Improved NSGA-III Algorithm. *MDPI Minerals* **2022**, *12*, 233. https://doi.org/10.3390/min12030233.
9. Cheng, Y.; Liu, S.; Wang, H. The Effects of Ore Blending on Comminution Behaviour and Product Quality. *Diva Portal* **2018**.
10. Fisher, T.; John, R.; Smith, M. Coal Blending Strategies Reduce Costs and Improve Quality in Mining. *Thermo Fisher* **2020**.
11. Khan, S.; Ali, M.; Li, X. Heuristic Strategies for Solving Complex Interacting Large-Scale Stockpile Blending Problems. *arXiv* **2022**.
12. Lindi, O.T.; Aladejare, A.E.; Ozoji, T.M.; Ranta, J.P. Uncertainty Quantification in Mineral Resource Estimation. *Natural Resources Research* **2024**, pp. 1–24.
13. Butcher, A.R.; Dehaine, Q.; Menzies, A.H.; Michaux, S.P. Characterisation of ore properties for geometallurgy. *Elements* **2023**, *19*, 352–358.
14. Poojari, M.; Vardhan, H.; Jathanna, H.M. Iron Ore Characterization Techniques in Mineral Processing. *Journal of The Institution of Engineers (India): Series D* **2024**, *105*, 543–551.
15. Gessner, K.; Blenkinsop, T.; Sorjonen-Ward, P. Characterization of ore-forming systems–advances and challenges. *Geological Society, London, Special Publications* **2018**, *453*, 1–6.
16. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G.; PRISMA Group*, t. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of internal medicine* **2009**, *151*, 264–269.
17. Bicak, O. A technique to determine ore variability in a sulphide ore. *Minerals Engineering* **2019**, *142*, 105927.
18. Mkurazhizha, H. The effects of ore blending on comminution behaviour and product quality in a grinding circuit-Svappavaara (LKAB) Case Study, 2018.
19. Li, J.; Liu, Y.; Zhao, X. Multi-Source and Multi-Target Iron Ore Blending Method in Open Pit Mining. *BIBLIOTEKA NAUKI* **2019**.
20. Xie, Y.; Neumann, A.; Neumann, F. Heuristic strategies for solving complex interacting large-scale stockpile blending problems. In Proceedings of the 2021 IEEE Congress on Evolutionary Computation (CEC). IEEE, 2021, pp. 1288–1295.
21. Xie, Y.; Neumann, A.; Neumann, F. Heuristic strategies for solving complex interacting stockpile blending problem with chance constraints. In Proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference, 2021, pp. 1079–1087.
22. Farghaly, M.G.; Ali, M.; Kim, J.G. Optimization of blending and production processes considering origin mines and metallurgical units using linear programming rules. *Geosystem Engineering* **2021**, *24*, 115–121. https://doi.org/10.1080/12269328.2021.1910574.
23. Ye, Z.; Hilden, M.M.; Yahyaei, M. A 3D cellular automata ore stockpile model – Part 2: Simulation and industrial validation of dynamic discharging and trajectory segregation mechanisms. *Minerals Engineering* **2023**, *200*, 108156. https://doi.org/10.1016/j.mineng.2023.108156.
24. Ye, Z.; Hilden, M.M.; Yahyaei, M. A 3D cellular automata ore stockpile model – Part 1: Simulation of size segregation. *Minerals Engineering* **2022**, *187*, 107816. https://doi.org/10.1016/j.mineng.2022.107816.

25. Servin, M.; Vesterlund, F.; Wallin, E. Digital Twins with Distributed Particle Simulation for Mine-to-Mill Material Tracking. *Minerals* **2021**, *11*, 524. https://doi.org/10.3390/min11050524.
26. Bowler, A.L.; Bakalis, S.; Watson, N.J. Monitoring mixing processes using ultrasonic sensors and machine learning. *Sensors* **2020**, *20*, 1813.
27. Benndorf, J.; Buxton, M.W.N. Sensor-based real-time resource model reconciliation for improved mine production control – a conceptual framework. *Mining Technology* **2016**, *125*, 54–64. https://doi.org/10.1080/14749009.2015.1107342.
28. Wambeke, T.; Benndorf, J. A Study of the Influence of Measurement Volume, Blending Ratios and Sensor Precision on Real-Time Reconciliation of Grade Control Models. *Mathematical Geosciences* **2018**, *50*, 801–826. https://doi.org/10.1007/s11004-018-9740-3.
29. Campos, P.H.A.; Costa, J.F.C.L.; Koppe, V.C.; Bassani, M.A.A.; Deutsch, C.V. Short-Term Schedule Optimization with Nonlinear Blending Models for Improved Metallurgical Recovery in Mining. *Mining, Metallurgy & Exploration* **2024**, pp. 1–15.
30. Nikkhah, A.; Vakylabad, A.B.; Hassanzadeh, A.; Niedoba, T.; Surowiak, A. An Evaluation on the Impact of Ore Fragmented by Blasting on Mining Performance. *Minerals* **2022**, *12*, 258. https://doi.org/10.3390/min12020258.
31. Qiu, P.; Pabst, T. Characterization of particle size segregation and heterogeneity along the slopes of a waste rock pile using image analysis. *Environmental Earth Sciences* **2023**, *82*, 573. https://doi.org/10.1007/s12665-023-11229-y.
32. Long, M.; Schafrik, S.; Kolapo, P.; Agioutantis, Z.; Sottile, J. Equipment and operations automation in mining: a review. *Machines* **2024**, *12*, 713.
33. Codoceo-Contreras, L.; Rybak, N.; Hassall, M. Exploring the impacts of automation in the mining industry: A systematic review using natural language processing. *Mining Technology* **2024**, *133*, 191–213.
34. Marques, D.M.; Costa, J.F.C. An algorithm to simulate ore grade variability in blending and homogenization piles. *International Journal of Mineral Processing* **2013**, *120*, 48–55.
35. Zhao, S.; Lu, T.F.; Koch, B.; Hurdsman, A. 3D stockpile modelling and quality calculation for continuous stockpile management. *International Journal of Mineral Processing* **2015**, *140*, 32–42. https://doi.org/10.1016/j.minpro.2015.04.012.
36. Wu, J.; Huang, L.; He, B.; Li, Z.; Chen, G.; Chen, G.; Li, X. Research on Multi-Objective Ore Blending Optimization Based on Non-Equilibrium Grade Polymetallic Mine of Shizhuyuan. *Minerals* **2022**, *12*, 1358. https://doi.org/10.3390/min12111358.
37. Reid, W.; Neumann, A.; Ratcliffe, S.; Neumann, F. Advanced ore mine optimisation under uncertainty using evolution. *arXiv preprint arXiv:2102.05235* **2021**.
38. Estay, R.; et al. Machine Learning Applications in Minerals Processing: A Review. *Minerals Engineering* **2018**, *123*, 70–84. https://doi.org/10.1016/j.mineng.2018.03.015.
39. Otsuki, A.; Jang, H. Prediction of particle size distribution of mill products using artificial neural networks. *ChemEngineering* **2022**, *6*, 92.
40. Both, C.; Dimitrakopoulos, R. Applied machine learning for geometallurgical throughput prediction—a case study using production data at the tropicana gold mining complex. *Minerals* **2021**, *11*, 1257.
41. WANG, L.g.; SONG, H.q.; BI, L.; CHEN, X. Optimization of open pit multielement ore blending based on goal programming. *Journal of Northeastern University (Natural Science)* **2017**, *38*, 1031.
42. Li, Z.; Guo, L.; Zhao, Y.; Peng, X.; Kyegyenbai, K. A Particle Size Distribution Model for Tailings in Mine Backfill. *Metals* **2022**, *12*, 594. https://doi.org/10.3390/met12040594.
43. Miriyala, S.S.; Mitra, K. Deep learning based system identification of industrial integrated grinding circuits. *Powder Technology* **2020**, *360*, 921–936.
44. Ghasemi, Z.; Neshat, M.; Aldrich, C.; Karageorgos, J.; Zanin, M.; Neumann, F.; Chen, L. A Hybrid Intelligent Framework for Maximising SAG Mill Throughput: An Integration of Expert Knowledge, Machine Learning and Evolutionary Algorithms for Parameter Optimisation. *arXiv preprint arXiv:2312.10992* **2023**.
45. Saldaña, M.; Gálvez, E.; Navarra, A.; Toro, N.; Cisternas, L.A. Optimization of the SAG grinding process using statistical analysis and machine learning: A case study of the Chilean copper mining industry. *Materials* **2023**, *16*, 3220.

46. Lindner, B.; Auret, L. Application of data-based process topology and feature extraction for fault diagnosis of an industrial platinum group metals concentrator plant. *IFAC-PapersOnLine* **2015**, *48*, 102–107.
47. Kumah, F.N.; Saim, A.K.; Oppong, M.N.; Arthur, C.K. Predicting open-pit mine production using machine learning techniques. *Journal of Sustainable Mining* **2024**, *23*, 118–131. https://doi.org/10.46873/2300-3960.1411.
48. Liu, B.; Zhang, D.; Gao, X. A Method of Ore Blending Based on the Quality of Beneficiation and Its Application in a Concentrator. *Applied Sciences* **2021**, *11*, 5092. https://doi.org/10.3390/app11115092.
49. Nakhaei, F.; Jovanović, I. Developments and applications of 3D imaging systems in mineral processing. *Journal of Mining and Metallurgy A: Mining* **2023**, *59*, 35–47.
50. Jung, D.; Choi, Y. Systematic Review of Machine Learning Applications in Mining: Exploration, Exploitation, and Reclamation. *Minerals* **2021**, *11*, 148. https://doi.org/10.3390/min11020148.
51. Liu, S.Q.; Liu, L.; Kozan, E.; Corry, P.; Masoud, M.; Chung, S.H.; Li, X. Machine learning for open-pit mining: a systematic review. *International Journal of Mining, Reclamation and Environment* **2025**, *39*, 1–39.
52. Statsenko, L.; Melkoumian, N.S. Modeling Blending Process at Open-Pit Stockyards: A Northern Kazakhstan Mining Company Case Study. In Proceedings of the Mine Planning and Equipment Selection; Drebenstedt, C.; Singhal, R., Eds., Cham, 2014; pp. 1017–1027. https://doi.org/10.1007/978-3-319-02678-7_98.
53. Parent, E.J.; Parent, S.É.; Parent, L.E. Determining soil particle-size distribution from infrared spectra using machine learning predictions: Methodology and modeling. *Plos one* **2021**, *16*, e0233242.
54. Shrivastava, S.; Deb, D.; Bhattacharjee, S. Prediction of particle size distribution curves of dump materials using convolutional neural networks. *Rock Mechanics and Rock Engineering* **2022**, pp. 1–9.

# Appendix B

## Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed

Submitted to the *International Journal of Mining Science and Technology*

**Matias Saavedra, Nathalie Risso, Moe Momayez, Ricardo Nunes, Victor Tenorio**

Department of Mining, Geological & Geophysical Engineering,
University of Arizona, 1235 James E. Rogers Way, Tucson, AZ 85719
United States

E-mail: matiassaavedra@arizona.edu
(520) 351 9225

# Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed

Matias Saavedra[1], Nathalie Risso[1*], Moe Momayez[1], Ricardo Nunes[1], Victor Tenorio[1]

[1*]Department of Mining and Geological Engineering, University of Arizona, 1235 E. James E. Rogers Way, Tucson, 85721, Arizona, USA.

*Corresponding author(s). E-mail(s): nrisso@arizona.edu;
Contributing authors: matiassaavedra@ariz ona.edu;
mmomayez@arizona.edu ; ricardonunes@arizona.edu ;
vtenorio@arizona.edu ;

**Abstract**

This study presents a data-driven case of study for modeling and reducing particle size distribution (PSD) variability at the feed of a semi-autogenous grinding (SAG) mill in a copper mining operation. Using two years of operational data from the stockpile and feeder system, Random Forest regression models, guided by unsupervised clustering, were developed to predict PSD outcomes based on feeder speeds, stockpile levels, and throughput. A web-based application was created to simulate and visualize PSD predictions under varying operating conditions. In addition to achieving high predictive accuracy (average $R^2$ exceeding 0.90), the work incorporates a univariate sensitivity analysis to quantify the influence of feeder setpoints on PSD targets, and a statistical assessment of variability, showing reduced dispersion in predicted PSD distributions when model-guided inputs are applied. These results demonstrate the potential of machine learning to support future decision-making strategies aimed at stabilizing SAG mill feed characteristics.

**Keywords:** PSD prediction, Stockpile management, Random Forest regression, Sensitivity, Variability

# 1 Introduction

Mining remains a cornerstone of global industrial development, supplying the raw materials essential for construction, energy, and advanced technologies [1]. As the global transition to renewable energy, electric vehicles, and digital infrastructure accelerates, demand for critical minerals like copper, lithium, and rare earths is rising. This trend pressures mining operations to boost throughput, reduce variability, and improve overall efficiency.

A key factor in achieving these improvements is the ability to manage particle size distribution (PSD), which directly influences grinding efficiency, flotation recovery, and overall plant stability. Particularly at the feed line of the grinding circuit (semi-autogenous grinding (SAG) mill) maintaining a consistent PSD is crucial for ensuring optimal energy utilization and downstream performance. However, PSD measurements are typically obtained through laboratory analyses, which limits the plant's ability to respond quickly to short-term operational variability [2].

Not all mining operations utilize stockpiles or the same blending systems. However, in many large-scale facilities, especially in copper mining, stockpiles serve as transitional buffers between ore extraction and processing stages. They help decouple upstream variability and enable basic blending strategies. In the case study considered, the copper mine operation employs a stockpile and feeder system to manage feed delivery to the SAG mill. This introduces complexity due to changes in ore type, moisture content, and feeder speed strategies, which in turn increase uncertainty and variability in feed characteristics. While modern sensor networks generate large volumes of high-frequency operational data, this information is often underutilized in predictive modeling or real-time decision support [3]. Leveraging these available data streams, recent advancements in sensor integration and centralized data infrastructures have enabled the application of data analysis and machine learning (ML) techniques, particularly regression-based models, to better understand and forecast complex process behaviors [4]. In mineral processing, such models have been successfully applied to predict grinding efficiency, flotation performance, and production bottlenecks with high accuracy [5–7].

Moreover, the dynamic behavior of stockpiles and material segregation during stacking has been studied through simulation tools such as the Discrete Element Method (DEM), revealing how changes in particle properties and stacking patterns lead to stratification and influence feed consistency [8]. Understanding and accounting for these effects are essential for improving the reliability of PSD predictions and optimizing blending strategies.

Enhancing SAG mill efficiency is often constrained by the variability of incoming ore, particularly fluctuations in the particle size distribution of the feed [9]. While equipment upgrades are costly and time intensive, many operations already possess extensive sensor infrastructure that may hold untapped potential for process optimization. This raises two central questions: *Can SAG mill efficiency be improved by leveraging existing infrastructure to identify and correlate the most relevant upstream operational variables that influence the PSD of the mill feed?* And more specifically, *Can a data-driven approach enable accurate characterization and prediction of PSD*

*based on sensor data from the stockpile and feeder system, thereby supporting better process control without costly equipment upgrades?*

This paper presents a case study from a copper mining operation that addresses these questions by predicting PSD at the SAG mill feed using historical operational data from the stockpile and feeder system. Cluster-specific Random Forest regression models were developed using two years of high frequency sensor data, with variables such as feeder speeds, stockpile levels, and ore throughput serving as key inputs. The study also includes a sensitivity analysis to quantify the effect of individual feeder adjustments on PSD targets, and a variability assessment comparing PSD distributions under historical versus model-adjusted scenarios. The results demonstrate strong predictive performance and highlight the potential of data-driven methods to stabilize and improve feed characteristics through informed operational adjustments [10].

# 2 Background

## 2.1 Stockpile Management and Particle Size Variability

In mineral processing, blending strategies are essential for homogenizing ore characteristics such as particle size, moisture content, and grade before feeding material into downstream circuits. Blending can be achieved through various operational configurations, including truck routing, dozer push techniques during stockpile construction, or controlled discharge via feeders. The choice of blending method depends on site-specific infrastructure, production scale, and material variability. Among these, stockpile-based blending using multiple feeders provides a dynamic and operationally flexible approach, especially when integrated with real-time process monitoring [11, 12].

Stockpiles are commonly used to store ore between the mine and the processing plant. However, during stacking, particle size segregation occurs due to differences in flowability and kinetic energy. Typically, finer particles concentrate near the center of the stockpile, while coarser material migrates toward the periphery [13], introducing significant variability into the feed. This variability can affect grinding efficiency, energy consumption, and downstream recovery processes [14, 15]. Although blending strategies and feeder adjustments are commonly used to mitigate these effects, they tend to be reactive and rely on limited sampling [16].

In mineral processing, it is well recognized that particle size distribution (PSD) tends to decrease progressively as material advances through the plant, especially following comminution stages [17]. Screening systems are often implemented after crushers or grinding circuits to control this progression, removing oversize particles and maintaining the PSD within a specified range. Nevertheless, the feed entering the primary grinding circuit (particularly the semi-autogenous grinding, or SAG, mill) can still exhibit substantial variability due to upstream inconsistencies.

This case study focuses on a copper mining operation that uses a stockpile and feeder system to deliver material to the SAG mill. Operational variables such as feeder speeds, stockpile levels, and throughput significantly influence the resulting PSD at the mill feed. Anticipating these feed characteristics requires predictive tools capable of utilizing high-frequency operational data. A schematic overview of the stockpile and feeder configuration is shown in Figure 1.

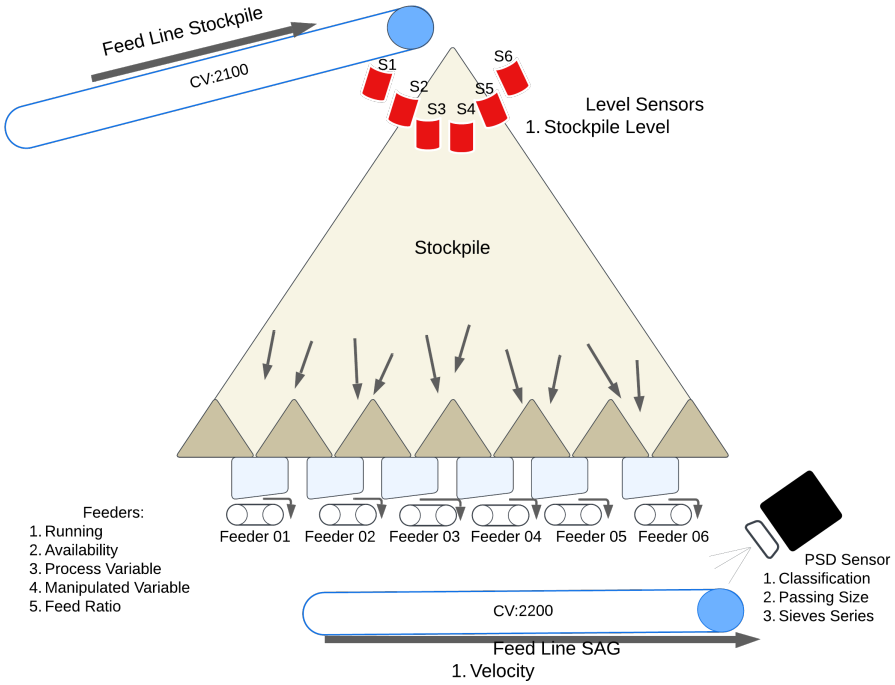


**Fig. 1**: Schematic overview of the stockpile and feeder system

Given PSD's critical role in determining SAG mill efficiency and overall plant performance, this study investigates the application of machine learning to predict PSD at the mill feed based on upstream operational data. In this case, PSD measurements were obtained using an image-based online analysis system installed on the conveyor belt feeding the SAG mill. This approach allows for high-frequency, non-intrusive monitoring of particle size characteristics. However, image-based systems typically require regular maintenance and calibration, and may not match the precision of laboratory-based measurements (particularly for fine particle size fractions). By leveraging two years of high-frequency sensor data along with image-derived PSD records, this study assesses the potential of data-driven models to enhance feed consistency and support future control strategies [18, 19].

## 2.2 Data Driven PSD Prediction

Traditional approaches like linear regression or ratio based blending often fail to model the nonlinear and multivariate nature of material flow through stockpiles. These methods assume fixed input output relationships, which limit their adaptability to varying ore types and dynamic process conditions [20].

Conversely, tree based machine learning algorithms, particularly Random Forest (RF) and Extreme Gradient Boosting (XGBoost), excel at capturing complex, nonlinear dependencies between inputs and PSD outcomes [21, 22]. These models have shown high accuracy in mining contexts, with $R^2$ values exceeding 0.95 [23], supporting their use in predicting PSD from stockpile and feeder system data.

Despite their performance, machine learning models are sometimes underutilized for decision support, mainly due to challenges like process variability and limited interpretability. To overcome these limitations, this study employs a hybrid approach combining unsupervised clustering with Random Forest regression. This enables the capture of regime specific behaviors in the stockpile feeder system. Beyond prediction, our case of study incorporates sensitivity analysis to evaluate input output relationships and a variability reduction component to assess the effect of model guided settings [24].

### 2.3 Sensitivity and Variability in Data Driven Modeling

While predictive models estimate expected outcomes for given input conditions, their practical utility increases when accompanied by an understanding of how those outcomes respond to input changes and vary under operational uncertainty [10, 25]. This is particularly relevant in mineral processing, where systems often exhibit strong nonlinearity and fluctuating regimes.

Two core concepts, *sensitivity* and *variability*, are used to characterize system behavior and guide control strategies. Sensitivity quantifies the influence of an input variable on the model output. In a fixed context, the local sensitivity of a predicted output $\hat{y}$ to input $x_i$ is given by:

$$S_i = \frac{\Delta \hat{y}}{\Delta x_i} \tag{1}$$

Here, $\Delta x_i$ is a deliberate change in the input, and $\Delta \hat{y}$ is the corresponding change in the predicted output. In our model, $x_i$ refers to feeder speed setpoints, and $\hat{y}$ denotes the predicted PSD values. Variability describes how predicted outputs disperse across a realistic range of input conditions [26, 27]. A common metric is the standard deviation of predictions:

$$\sigma_{\hat{y}} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \bar{y})^2} \tag{2}$$

where $\hat{y}_i$ is the PSD prediction for the $i$-th input scenario and $\bar{y}$ is the mean prediction across $N$ samples. This measure reflects the model's robustness to natural variation in operational conditions.

## 3 Methodology

### 3.1 Dataset and Preprocessing

The dataset used in this study was obtained from the data infrastructure of a copper mining operation, with a focus on the stockpile and feeder system responsible

for delivering ore to the primary SAG mill. It consists of two years of high-frequency operational data collected at five-minute intervals, yielding over 210,000 synchronized records. This data integrates both manipulated variables (MVs) and process variables (PVs), offering a comprehensive representation of upstream plant conditions. Key variables include feeder-related information such as operating status, availability, and both reference (MV) and measured (PV) feeder velocities. Descriptive statistics for these variables are summarized in Table 1, highlighting their distributions, variability, and data quality (e.g., missing values and out-of-range detections). In addition to the feeder system, the dataset encompasses operational indicators related to stockpile levels, conveyor belt velocities, material throughput, and SAG mill conditions, including power draw, rotational speed, solids concentration, and load weight. Table 2 (continued) presents statistical summaries for these remaining operational variables, which together serve as inputs for characterizing and predicting the particle size distribution (PSD) of SAG mill feed.

**Table 1**: Descriptive statistics of feeder-related variables.

| Group | Min | Max | Mean | Variability | NaN | Out-of-Range |
|---|---|---|---|---|---|---|
| Feeder 01 Ratio | 0.018 | 1 | 0.975 | 1.007 | 53 | 3135 |
| Feeder 02 Ratio | 0.3 | 1 | 0.983 | 0.712 | 53 | 2567 |
| Feeder 03 Ratio | 0.3 | 1 | 0.956 | 0.733 | 53 | 1373 |
| Feeder 04 Ratio | 0.3 | 1 | 0.978 | 0.716 | 53 | 2624 |
| Feeder 05 Ratio | 0.3 | 1 | 0.982 | 0.713 | 53 | 2737 |
| Feeder 06 Ratio | 0.3 | 1 | 0.972 | 0.720 | 53 | 2423 |
| Feeder 01 Running Status | 0 | 1 | 0.782 | 1.279 | 0 | 0 |
| Feeder 02 Running Status | 0 | 1 | 0.831 | 1.203 | 0 | 0 |
| Feeder 03 Running Status | 0 | 1 | 0.752 | 1.330 | 0 | 0 |
| Feeder 04 Running Status | 0 | 1 | 0.807 | 1.239 | 0 | 0 |
| Feeder 05 Running Status | 0 | 1 | 0.810 | 1.234 | 0 | 0 |
| Feeder 06 Running Status | 0 | 1 | 0.818 | 1.222 | 0 | 0 |
| Feeder 01 Availability | 0 | 1 | 0.920 | 1.087 | 0 | 0 |
| Feeder 02 Availability | 0 | 1 | 0.953 | 1.049 | 0 | 2976 |
| Feeder 03 Availability | 0 | 1 | 0.987 | 1.013 | 0 | 831 |
| Feeder 04 Availability | 0 | 1 | 0.954 | 1.048 | 0 | 2902 |
| Feeder 05 Availability | 0 | 1 | 0.909 | 1.100 | 0 | 0 |
| Feeder 06 Availability | 0 | 1 | 0.958 | 1.044 | 0 | 2653 |
| Feeder 01 Velocity MV | 0 | 437.225 | 43.813 | 9.979 | 54 | 3158 |
| Feeder 02 Velocity MV | 0 | 274.097 | 44.324 | 6.184 | 54 | 3158 |
| Feeder 03 Velocity MV | 0 | 437.225 | 42.558 | 10.274 | 54 | 3158 |
| Feeder 04 Velocity MV | 0 | 437.225 | 44.139 | 9.906 | 54 | 3158 |
| Feeder 05 Velocity MV | 0 | 437.225 | 44.125 | 9.909 | 54 | 3158 |
| Feeder 06 Velocity MV | 0 | 437.225 | 43.675 | 10.011 | 54 | 3158 |
| Feeder 01 Velocity PV | 0 | 4.936 | 1.834 | 2.691 | 53 | 6316 |
| Feeder 02 Velocity PV | 0 | 4.967 | 1.940 | 2.560 | 53 | 6316 |
| Feeder 03 Velocity PV | -0.001 | 5.049 | 1.723 | 2.930 | 53 | 6316 |
| Feeder 04 Velocity PV | 0 | 4.999 | 1.892 | 2.642 | 53 | 6316 |
| Feeder 05 Velocity PV | 0 | 5.052 | 1.897 | 2.663 | 53 | 6316 |
| Feeder 06 Velocity PV | 0.001 | 4.985 | 1.874 | 2.659 | 53 | 6316 |

Particle size distribution (PSD) data were obtained from an on site Split system, which reported percent passing values for F10 through F90 and TOPSIZE, as well as full sieve measurements. Bulk size classifications, fines, intermediate, and coarse percentages, were also reported by the measurement system implemented on site. Together, these variables capture ore transport dynamics, blending behavior, equipment activity, and feed characteristics. The preprocessing workflow was as follows:

- **Handling Missing Values:** Non-informative tags and calculated columns were excluded. Rows containing missing values were removed to ensure training consistency and avoid artifacts.
- **Outlier Removal:** Extreme values were filtered using the inter-quartile range (IQR) method. This retained approximately 84% of the original data while preserving meaningful variability.

**Table 2**: Descriptive statistics of remaining operational variables

| Group | Min | Max | Mean | Variability | NaN | Out-of-Range |
|---|---|---|---|---|---|---|
| Stockpile Level 01 | 0 | 98.516 | 71.545 | 1.377 | 596 | 6262 |
| Stockpile Level 02 | 0.001 | 99.145 | 70.883 | 1.399 | 53 | 6316 |
| Stockpile Level 03 | 27.135 | 103.645 | 80.763 | 0.947 | 53 | 3777 |
| Stockpile Level 04 | 0.001 | 98.931 | 70.421 | 1.405 | 53 | 6316 |
| Stockpile Level 05 | 0.202 | 100.984 | 72.171 | 1.396 | 53 | 6316 |
| Stockpile Level 06 | 32.723 | 101.874 | 76.157 | 0.908 | 53 | 6316 |
| Coarse Fraction (%) | 0 | 99.885 | 59.805 | 1.670 | 53 | 6282 |
| Intermediate Fraction (%) | 0 | 29.872 | 18.765 | 1.592 | 53 | 6316 |
| Fine Fraction (%) | 0 | 93.431 | 18.914 | 4.940 | 53 | 6006 |
| F10 | 0 | 6.556 | 0.038 | 173.499 | 53 | 6316 |
| F20 | 0 | 17.220 | 0.169 | 101.676 | 53 | 6316 |
| F30 | 0 | 17.297 | 0.413 | 41.858 | 53 | 6316 |
| F40 | 0 | 17.329 | 0.802 | 21.612 | 53 | 6316 |
| F50 | 0 | 18.188 | 1.344 | 13.533 | 53 | 6316 |
| F60 | 0 | 21.300 | 2.001 | 10.647 | 53 | 6316 |
| F70 | 0 | 22.018 | 2.796 | 7.875 | 53 | 6316 |
| F80 | 0 | 22.559 | 3.778 | 5.971 | 53 | 6316 |
| F90 | 0 | 23.088 | 5.123 | 4.506 | 53 | 6316 |
| Ftopsize | 0 | 23.452 | 7.869 | 2.980 | 53 | 6316 |
| PSD Sieve Series 1 | 0 | 90.447 | 37.138 | 2.435 | 53 | 6316 |
| PSD Sieve Series 2 | 0 | 98.009 | 47.415 | 2.067 | 53 | 6316 |
| PSD Sieve Series 3 | 0 | 100.000 | 95.713 | 1.045 | 53 | 3158 |
| PSD Sieve Series 4 | 0 | 99.486 | 53.834 | 1.848 | 53 | 6316 |
| PSD Sieve Series 5 | 0 | 99.925 | 59.801 | 1.671 | 53 | 6275 |
| PSD Sieve Series 6 | 0 | 100.000 | 65.199 | 1.534 | 53 | 6253 |
| PSD Sieve Series 7 | 0 | 100.000 | 70.090 | 1.427 | 53 | 6316 |
| PSD Sieve Series 8 | 0 | 100.000 | 74.543 | 1.342 | 53 | 6316 |
| PSD Sieve Series 9 | 0 | 100.000 | 78.564 | 1.273 | 53 | 6316 |
| PSD Sieve Series 10 | 0 | 100.000 | 85.365 | 1.171 | 53 | 6316 |
| PSD Sieve Series 11 | 0 | 100.000 | 90.319 | 1.107 | 53 | 6316 |
| PSD Sieve Series 12 | 0 | 100.000 | 93.447 | 1.070 | 53 | 6132 |
| PSD Sieve Series 13 | 0 | 100.000 | 94.988 | 1.053 | 53 | 6315 |
| PSD Sieve Series 14 | 0 | 100.000 | 95.491 | 1.047 | 53 | 3107 |

- **Nonlinear Correlation Matrix:** Distance correlation analysis was applied to detect and remove highly redundant variables. Features with a correlation above 0.95 were excluded.
- **Feature Importance:** A baseline Random Forest regression model was employed to assess the relative importance of upstream operational variables in predicting the PSD of the SAG mill feed. This data-driven approach enabled the identification of the most influential factors across a high-dimensional dataset, without imposing assumptions about linearity or independence among features.
  Feature importance was calculated using the mean decrease in impurity criterion, which quantifies how much each variable contributes to reducing prediction error. The ranking results, shown in Figure 2, revealed that the most relevant inputs were primarily related to feeder velocities, both manipulated (MV) and measured (PV), stockpile levels, and the material throughput to the SAG mill.

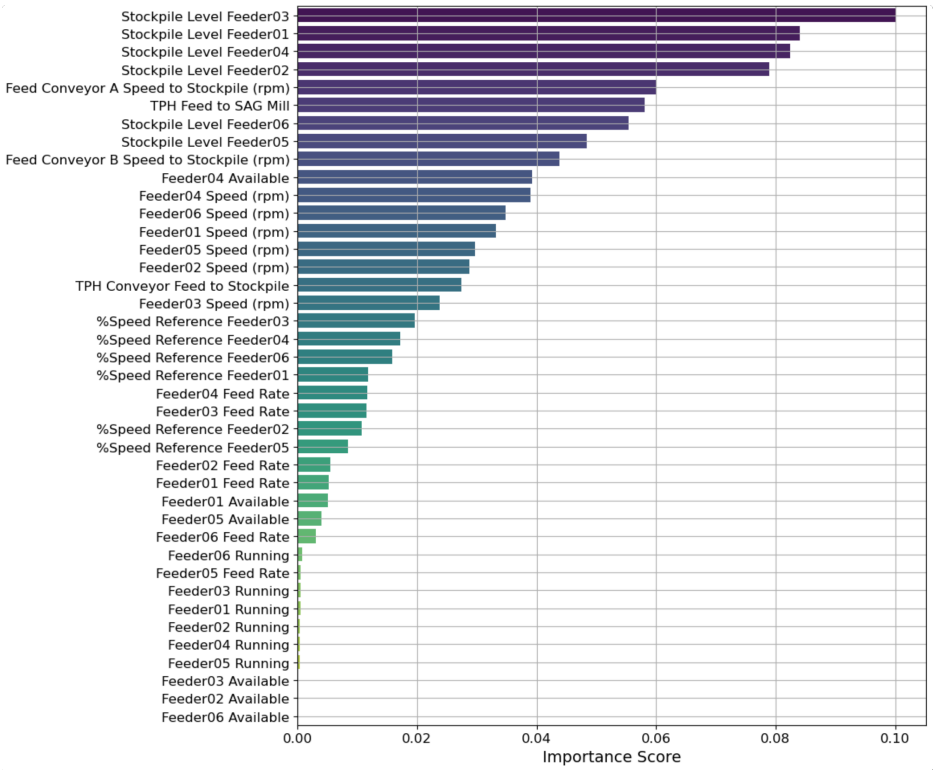


**Fig. 2**: Feature importance ranking.

The prominence of these variables provides key insights into operational dynamics. Feeder speeds directly control the proportion and sequence in which ore is drawn from different parts of the stockpile, affecting the blending of materials with distinct particle size profiles. Stockpile levels, in turn, reflect the spatial distribution and vertical stratification of material within the pile, conditions that strongly influence the size distribution of the discharged ore due to segregation during stacking. Finally, throughput variables capture the real-time intensity of material transfer, which modulates how quickly changes in stockpile composition manifest at the mill feed.

These insights reinforce the idea that effective control over feed particle size does not depend on a single sensor or action, but rather on a coordinated management of the stockpile, feeder system. By selecting the top 25 variables, those with the strongest influence on PSD, for downstream modeling, we focus the prediction task on the levers most likely to offer actionable control, enabling more informed decision-making in daily operations and long-term planning.

- **Dimensionality Reduction via PCA:** Principal Component Analysis (PCA) was performed on the selected features to reduce dimensionality before clustering.

Components explaining over 95% of the total variance were retained to minimize noise and computational complexity.

## 3.2 Clustering Analysis

Unsupervised clustering was applied to segment the dataset into distinct operational regions based on feeding conditions and PSD (Particle Size Distribution) behavior. This approach enables the development of specialized models tailored to different ore types and blending strategies, ultimately enhancing prediction accuracy and interpretability.

The PCA-transformed space improved variable separability and reduced noise, which facilitated more effective clustering. The optimal number of clusters was determined using the elbow method, which assesses the trade-off between the number of clusters and the within-cluster variance (inertia). A value of $K = 3$ was selected as it provided a favorable balance—further increases offered only marginal improvements and risked over-segmentation.

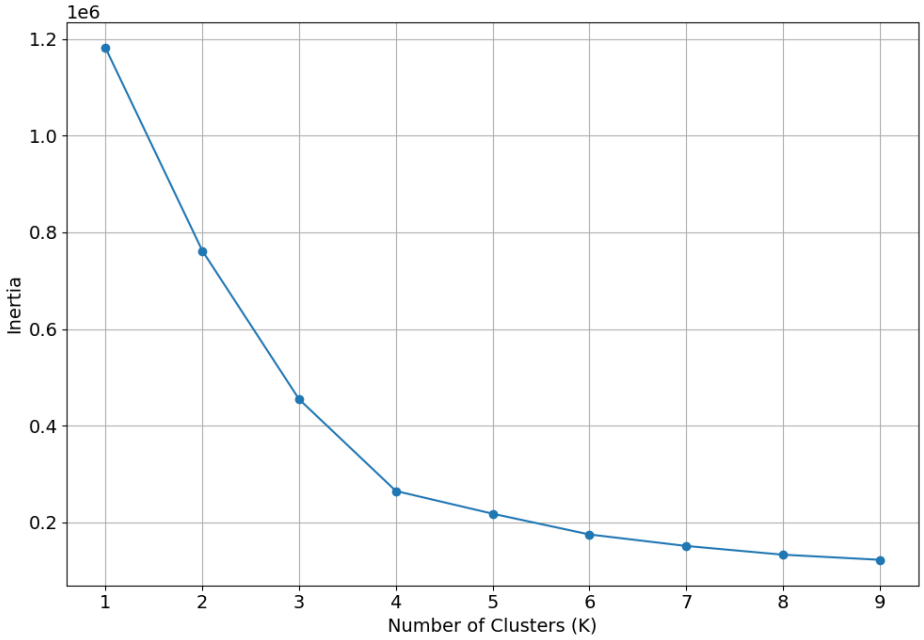


**Fig. 3**: Elbow method showing the optimal number of clusters at $K = 3$.

Beyond statistical criteria, domain knowledge further supports the choice of three clusters. In SAG mill operations, the particle size distribution of the feed is a critical driver of downstream performance, and it is often qualitatively categorized into three broad classes: fine, intermediate, and coarse. These categories align well with observed

blending and feeding dynamics at the site, which are influenced by factors such as stockpile geometry, feeder behavior, and ore fragmentation patterns. By aligning the clustering structure with these operational realities, the $K = 3$ configuration not only captures the statistical heterogeneity of the data but also yields meaningful, actionable groupings that can be linked to distinct operational modes and control strategies. Once $K = 3$ was selected, the K-Means algorithm was applied to the PCA-transformed data. Each data point was assigned a cluster label representing an operational regime characterized by similar feeder speeds, stockpile levels, and PSD patterns.

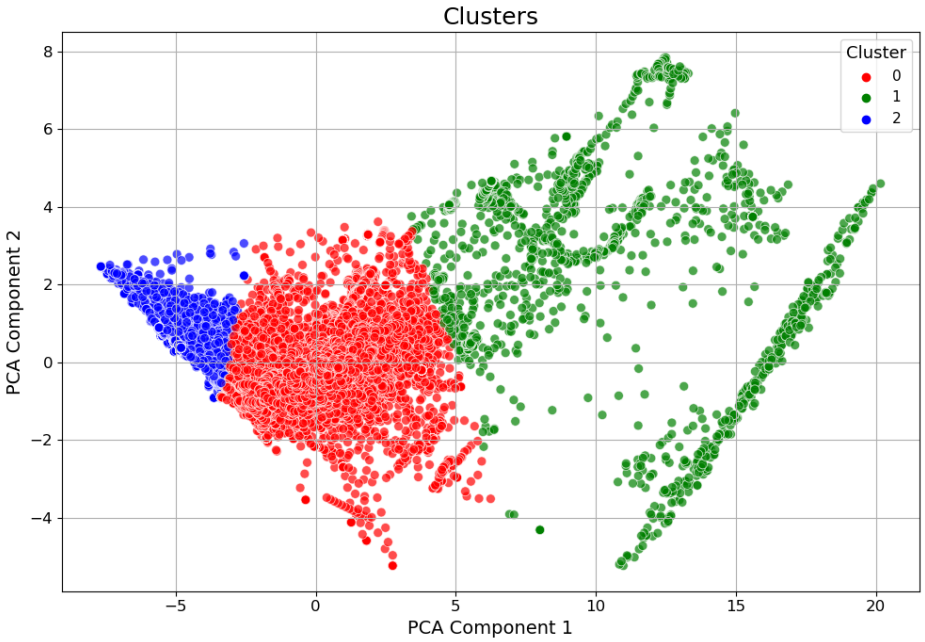


**Fig. 4**: Visualization of K-Means clustering results in 2D PCA space.

These cluster labels were used throughout the modeling pipeline to guide the training of cluster-specific Random Forest models and evaluate prediction performance under regime-dependent conditions.

## 3.3 Predictive Modeling

Separate Random Forest regression models were trained for each cluster to predict five PSD targets: F10, F30, F50, F70, and TOPSIZE. Each cluster's data was randomly split into 80% for training and 20% for testing.

To maintain consistency and facilitate comparison, default hyperparameter were used. Specifically, Random Forest was configured with *n_estimators* set to 100

and *random_state* to 42. Model performance was evaluated using the coefficient of determination ($R^2$) and Root Mean Square Error (RMSE) for each PSD target.

Figure 5 illustrates an example of model performance for the F30 prediction, showing the close agreement between observed and predicted values.

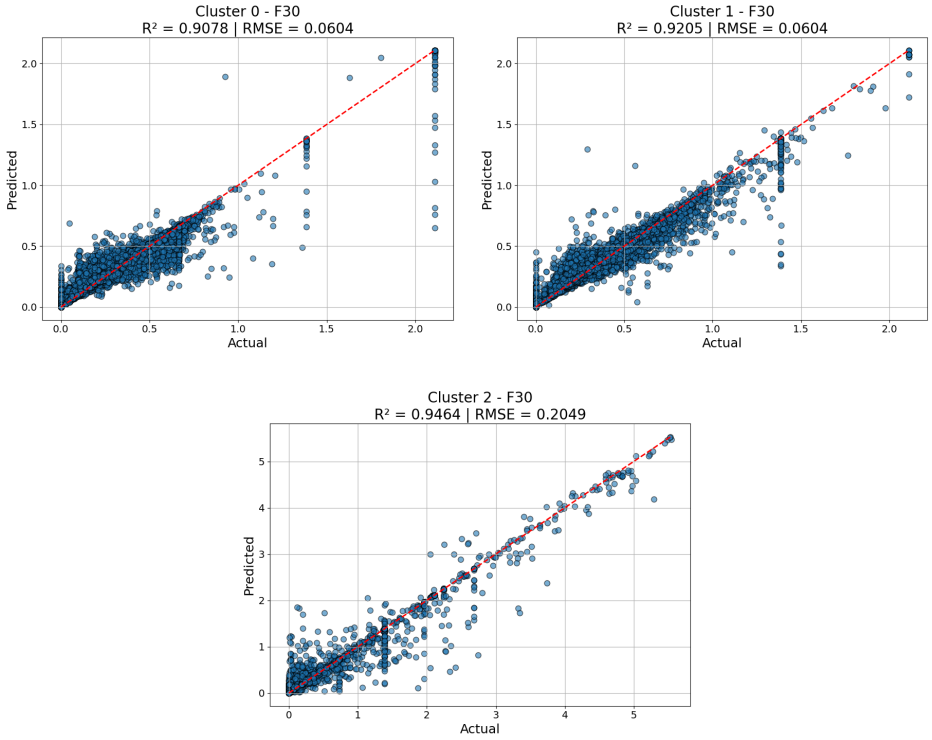


**Fig. 5**: Observed vs. predicted values for F30 using the Random Forest regression model across three clusters.

A web-based application was developed using Streamlit to facilitate model visualization and operator interaction. Users can adjust feeder speed setpoints, monitor current stockpile and throughput conditions, and visualize the predicted PSD curves. The interface dynamically generates tabular summaries and plots for scenario exploration and decision support.

# 4 Results

This section presents the main findings from the predictive modeling and analysis, which combines unsupervised clustering with supervised regression to capture the complex and nonlinear behavior of the SAG mill feed system. The hybrid approach, which integrates cluster-based segmentation with Random Forest models trained for each cluster, is well suited to handle the high variability and multi modal characteristics of the dataset. By grouping data into operational regimes defined by distinct feeding conditions and particle size distribution (PSD) patterns, the model can specialize its predictions based on context, which improves accuracy and interpretability.

This strategy offers several advantages. Clustering enhances model precision by isolating patterns associated with different ore types or blending scenarios, while regression enables quantification of the relationship between upstream variables and PSD targets within each regime. The results confirm the effectiveness of this approach, as each cluster-specific Random Forest model achieved high predictive performance when estimating PSD metrics based on operational variables.

The remainder of this section is structured as follows. First, we evaluate the accuracy of the cluster-specific models in predicting PSD targets. Next, a sensitivity analysis quantifies the influence of individual feeder setpoints on PSD outcomes, providing actionable insights into controllability. Finally, we compare the statistical distributions of predicted PSD values under baseline and model-adjusted conditions in order to assess the potential for reducing variability through optimized control.

## 4.1 Model Prediction Accuracy

Cluster-specific Random Forest regression models were trained to predict five PSD targets, F10, F30, F50, F70, and TOPSIZE, using upstream operational variables. Feature selection was guided by importance rankings derived from prior analysis, highlighting feeder speeds, stockpile levels, and throughput indicators.

To evaluate the trade-off between model complexity and predictive performance, three variable subsets were tested: the top 4, 17, and 25 variables. Performance was assessed using the coefficient of determination ($R^2$) and Root Mean Square Error (RMSE, in inches). As shown in Table 3, the 17-variable model delivered the highest average $R^2$ across all clusters, with Cluster 0 achieving $R^2 = 0.948$ and RMSE = 0.511 in. This configuration offered a compact yet highly accurate representation of the system.

Despite the 17-variable model yielding the best average accuracy, we selected the 25-variable configuration for further development and integration into the control interface. This decision was driven by the inclusion of all *manipulated variables (MVs)*, such as feeder speed setpoints, which are directly adjustable by operators or optimization systems.

Including MVs ensures that the model is not only predictive but also *actionable*. That is, it captures the cause-effect relationships between operational inputs and PSD outcomes, allowing for practical use in decision support or automatic control systems. The 25-variable model thus balances statistical performance with operational relevance, preserving control flexibility while maintaining high predictive power.

**Table 3**: Average $R^2$ and RMSE (in inches) across clusters for different variable selections

| Variable Set | Cluster | Avg $R^2$ | Avg RMSE (in) | Remarks |
|---|---|---|---|---|
| Top 4 variables | 0 | 0.431 | 0.528 | Limited predictive scope |
| | 1 | 0.477 | 0.653 | High RMSE, underfitting |
| | 2 | 0.937 | 0.297 | Acceptable in low-variance zone |
| Top 17 variables | 0 | **0.948** | **0.511** | Best overall accuracy |
| | 1 | 0.889 | 0.208 | Balanced performance |
| | 2 | 0.922 | 0.186 | Low error, good fit |
| Top 25 variables (incl. MVs) | 0 | 0.900 | 0.197 | Slight drop in $R^2$, lower RMSE |
| | 1 | **0.922** | **0.188** | Strong control fidelity |
| | 2 | **0.922** | 0.635 | Outlier due to target variance |

Although the $R^2$ value in Cluster 0 decreased by about *5.1%* compared to the 17-variable case (from 0.948 to 0.900), the RMSE improved slightly (from 0.511 in to 0.197 in), and in Clusters 1 and 2 the performance remained comparable or superior. This trade-off supports the development of interpretable and controllable models essential for process optimization and real-time SAG mill feed management.

As shown in Table 4, final models using the full 25-variable input space performed robustly across all clusters and targets. Cluster 1 exhibited the strongest performance, with $R^2$ values exceeding 0.90 for all PSD fractions. Conversely, Cluster 2—associated with more erratic feed conditions, demonstrated lower performance, particularly for coarser fractions, which may reflect greater variability or unmodeled dynamics.

## 4.2 Sensitivity Analysis

To quantify the impact of individual feeder setpoints on particle size distribution (PSD), a univariate sensitivity analysis was conducted. Each manipulated variable (MV), corresponding to one of the six calculated feeder speed references, was systematically varied across its historical operating range using 50 evenly spaced values. All remaining input variables were held constant at their baseline (mean) values to create a consistent input scenario.

For each generated scenario, the operational cluster was dynamically determined based on the mean value of the corresponding PSD target. The appropriate cluster-specific Random Forest model, along with its associated scaler and Principal Component Analysis (PCA) transformer, was used to pre process the inputs and predict PSD outcomes. This process was repeated independently for each PSD target (F10, F30, F50, F70, and TOPSIZE), generating a series of response curves.

Two sensitivity metrics were computed to quantify the influence of each feeder on PSD targets:

**Table 4**: Detailed regression metrics ($R^2$ and RMSE in inches) per cluster and PSD target using a 60 minute rolling window.

| Cluster | Target | $R^2$ | RMSE (in) |
|---|---|---|---|
| 0 | F10 | 0.9057 | 0.0038 |
| 0 | F30 | 0.9205 | 0.0590 |
| 0 | F50 | 0.9027 | 0.1750 |
| 0 | F70 | 0.8868 | 0.2515 |
| 0 | TOPSIZE | 0.8846 | 0.3566 |
| 1 | F10 | 0.9373 | 0.0149 |
| 1 | F30 | 0.9505 | 0.2414 |
| 1 | F50 | 0.9112 | 0.5928 |
| 1 | F70 | 0.9450 | 0.5374 |
| 1 | TOPSIZE | 0.9826 | 0.4048 |
| 2 | F10 | 0.7956 | 0.0012 |
| 2 | F30 | 0.8039 | 0.0198 |
| 2 | F50 | 0.8108 | 0.0760 |
| 2 | F70 | 0.8466 | 0.1808 |

- **Absolute Delta** ($\Delta$)**:** The difference between the maximum and minimum predicted values across the MV sweep.
- **Normalized Slope:** The slope of the response curve, normalized by the historical range of the respective PSD variable. This indicates the relative responsiveness of each PSD target to changes in feeder speed.

The analysis identified *Feeders 2, 3, and 5* as having the strongest influence across both fine and coarse size fractions. Adjustments to their setpoints produced measurable changes in predicted PSD outputs. For example, Feeder 2 exhibited a sensitivity of approximately *0.0052 inches* for F50 and *0.0028 inches* for F70, with Feeder 5 showing comparable magnitudes.

*Feeder 4* displayed a distinct non-monotonic response: increasing its setpoint led to lower predicted values for finer fractions (F10–F50), yet simultaneously raised predictions for coarser targets, particularly F70 (*0.0076 inches*) and TOPSIZE (*0.0099 inches*). This asymmetric effect suggests a complex interaction likely driven by internal stockpile segregation patterns and ore flow dynamics.

Table 5 summarizes the normalized sensitivity slopes for each MV and PSD combination, illustrating the relative control potential of each feeder.

**Table 5**: Percentage sensitivity of PSD targets (% of range) per 1% change in each feeder MV.

| Feeder MV | F10 (%) | F30 (%) | F50 (%) | F70 (%) | TOPSIZE (%) |
|---|---|---|---|---|---|
| WYC20561.MV | 11.28 | 6.11 | 4.39 | 5.24 | 4.65 |
| WYC20562.MV | **16.78** | 8.85 | 4.71 | 6.80 | 5.07 |
| WYC20563.MV | 16.75 | 8.80 | 4.88 | 6.81 | 5.16 |
| WYC20564.MV | 11.58 | 6.85 | **7.06** | **7.46** | **7.08** |
| WYC20565.MV | 16.77 | **8.84** | 5.22 | 6.79 | 5.07 |

Overall, the sensitivity analysis identified which feeders exert the strongest influence on PSD targets, offering insights for the development of future control strategies and operational tuning aimed at managing feed characteristics to the SAG mill.

### 4.3 Variability Analysis

A critical aspect of operational control is not only predicting expected process outcomes, but also reducing their variability. In the context of SAG mill feed management, fluctuations in particle size distribution (PSD) can lead to inconsistent grinding behavior, suboptimal energy efficiency, and diminished metal recovery. Therefore, this study also evaluates how effectively the proposed predictive model can support more consistent PSD outputs under model-guided operating conditions.

The variability analysis was conducted by comparing the statistical distributions of PSD values under two scenarios: historical baseline values and predicted values generated through systematic feeder adjustment. These predictions were obtained as part of the sensitivity analysis, where each manipulated variable (MV) was varied independently across its historical range while all other inputs were fixed at mean values. For each scenario, PSD predictions were generated using the appropriate cluster-specific Random Forest model, PCA transformer, and scaler. The resulting datasets were then cleaned to remove incomplete or invalid records, ensuring fair statistical comparison.

Gaussian distribution fits were applied to both original and predicted values for five PSD targets (F10, F20, F30, F50, and TOPSIZE) to visualize and evaluate variability. As shown in Figure 6, the predicted distributions are visibly narrower and more centered around their means, especially for coarser size fractions. This pattern suggests a reduction in random noise and a more stable output when operational decisions are informed by the model. For clarity and cosmetic design in the paper, only the original distribution of F80 and the predicted distribution of F80 are shown in the figure.

To quantify this effect, Table 6 reports the mean ($\mu$) and standard deviation ($\sigma$) for each PSD target before and after model-based adjustment, along with the corresponding percentage reduction in variability. The results indicate a reduction of up to 15.2% in standard deviation for TOPSIZE, with consistent improvements observed across F10 through F70. These findings highlight the ability of the proposed application not only to predict PSD accurately, but also to support operational strategies aimed at enhancing feed consistency.

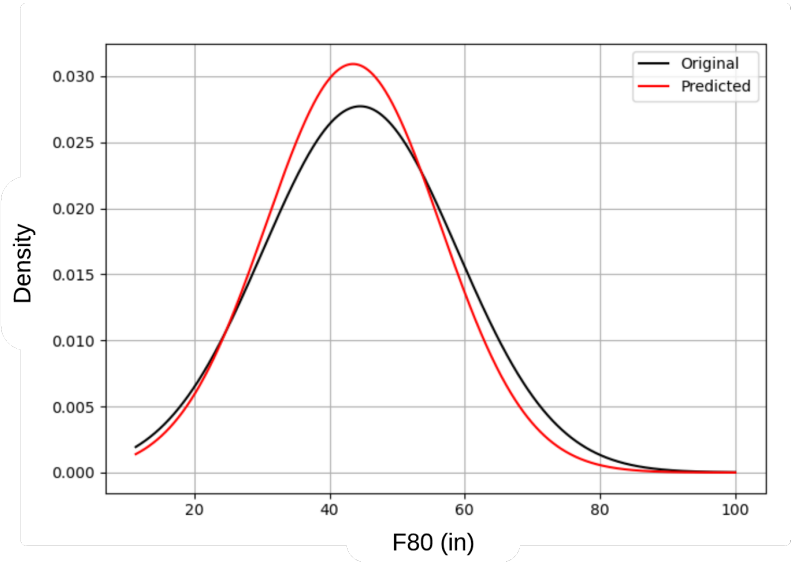


**Fig. 6**: Gaussian distribution fits for original and predicted PSD values.

**Table 6**: Comparison of original vs predicted PSD distributions and variability reduction.

| Variable | Original $\mu$ | Original $\sigma$ | Predicted $\mu$ | Predicted $\sigma$ | Reduction (%) |
|---|---|---|---|---|---|
| F10 | 0.0221 | 0.0207 | 0.0221 | 0.0179 | 13.5% |
| F30 | 0.3638 | 0.3618 | 0.3637 | 0.3240 | 10.5% |
| F50 | 1.3270 | 1.0255 | 1.3268 | 0.9004 | 12.2% |
| F70 | 2.8044 | 1.5133 | 2.8040 | 1.3411 | 11.4% |
| TOPSIZE | 7.4259 | 2.1170 | 7.4250 | 1.7948 | 15.2% |

## 5 Conclusion

Addressing the challenge of SAG mill efficiency constrained by feed PSD variability, this study demonstrated that particle size distribution at the SAG mill feed can be accurately predicted using operational data from the stockpile and feeder system. Leveraging two years of high-frequency historical data, a data-driven steup was developed based on Random Forest regression and unsupervised clustering, effectively capturing nonlinear dependencies and regime-specific patterns. The models achieved strong predictive performance, with $R^2$ values exceeding 90% for most PSD targets.

In addition to prediction, a univariate sensitivity analysis was conducted to quantify the impact of individual feeder setpoints on PSD outcomes. Results indicated that coarser size fractions, such as F70 and TOPSIZE, exhibit higher sensitivity to feeder adjustments, whereas finer fractions show more limited responsiveness. These

findings provide actionable insights for feed control, highlighting which variables are most influential for adjusting PSD behavior.

This case of study further demonstrated the potential to reduce PSD variability through model guided operating strategies. By comparing the standard deviation of predicted versus original PSD values under fixed operational inputs, the study showed that targeted feeder adjustments can narrow particle size distributions, especially for coarse fractions, resulting in more stable SAG mill feed conditions.

All components of the study were integrated into an interactive web application, enabling operators to simulate PSD responses and evaluate input output relationships under various scenarios. This decision support tool offers practical value for operational teams and establishes a foundation for future implementation in environments with real time data availability.

# 6 Future Work

Future work can focus on deploying the developed web application in a real operational environment by establishing a live connection to the site's data infrastructure. This implementation will enable continuous PSD prediction based on real time operational data and allow for practical validation of model performance under dynamic plant conditions. Operational feedback will provide essential insights into the robustness and usability of the tool in real time decision making contexts.

Future development will also explore the integration of more advanced machine learning models, such as Convolutional Neural Networks (CNNs), to capture spatiotemporal patterns and transient effects in the data. These architectures have the potential to improve prediction accuracy in rapidly changing operational regimes or during shifts in ore characteristics.

In addition, the study can be extended to incorporate a formal controllability analysis for nonlinear systems. This includes evaluating the system's ability to drive PSD outcomes toward desired targets under constraints, using data driven techniques suited for nonlinear and time varying environments. This step will help quantify the practical maneuverability of the stockpile and feeder system beyond uni variate sensitivity, laying the groundwork for future control strategy development.

Ultimately, the integration of real time prediction, advanced modeling, and controllability analysis will support the development of feedforward or model predictive control (MPC) architectures. These systems will enable autonomous and adaptive SAG mill control, contributing to greater stability, energy efficiency, and process optimization across mineral processing operations.

# References

[1] Ali, S.H., Giurco, D., Arndt, N., Nickless, E., Brown, G., Demetriades, A., Durrheim, R., Enriquez, M.A., Kinnaird, J., Littleboy, A., *et al.*: Mineral supply for sustainable development requires resource governance. Nature **543**(7645), 367–372 (2017)

[2] Maass, S.: Real-time particle-size analysis technologies: Overview and case study. Chemical Engineering **130**(7) (2023)

[3] Hasidi, O., Abdelwahed, E.H., El Alaoui-Chrifi, M.A., Qazdar, A., Benzakour, I., Bourzeix, F., Bendaouia, A.: Digital twin of minerals processing operations for an advanced monitoring and supervision: froth flotation process case study. The International Journal of Advanced Manufacturing Technology **132**(1), 1031–1049 (2024)

[4] McCoy, J.T., Auret, L.: Machine learning applications in minerals processing: A review. Minerals Engineering **132**, 95–109 (2019)

[5] Moraga, C., Astudillo, C.A., Estay, R., Maranek, A.: Enhancing comminution process modeling in mineral processing: A conjoint analysis approach for implementing neural networks with limited data. Mining **4**(4), 966–982 (2024)

[6] Pural, Y.E., Ledezma, T., Hilden, M., Forbes, G., Boylu, F., Yahyaei, M.: Application of machine learning for generic mill liner wear prediction in semi-autogenous grinding (sag) mills. Minerals **14**(12), 1200 (2024)

[7] Lopez, P., Reyes, I., Risso, N., Momayez, M., Zhang, J.: Machine learning algorithms for semi-autogenous grinding mill operational regions' identification. Minerals **13**(11), 1360 (2023)

[8] Gómez, R., Skrzypkowski, K., Moncada, M., Castro, R., Lazo, R.: Segregation modeling in stockpile using discrete element method. Applied Sciences **12**(23), 12449 (2022)

[9] Razani, M., Masoumi, A., Rezaeizadeh, M., Noaparast, M.: Prediction model to estimate the specific energy and product particle size distribution of semi-autogenous mill. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering **233**(2), 306–313 (2019)

[10] Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S.: Global Sensitivity Analysis: The Primer. John Wiley & Sons, Chichester (2008)

[11] Mkurazhizha, H.: The effects of ore blending on comminution behaviour and product quality in a grinding circuit-Svappavaara (LKAB) Case Study (2018)

[12] Robinson, G.: How much would a blending stockpile reduce variation? Chemometrics and intelligent laboratory systems **74**(1), 121–133 (2004)

[13] Ye, Z., Yahyaei, M., Hilden, M., Powell, M.: Novel size segregation indices for multi-sized particle stockpiles. Minerals Engineering **201**, 108165 (2023)

[14] Razani, M., Masoumi, A., Rezaeizadeh, M., Noaparast, M.: Evaluating the effect

of feed particles size and their hardness on the particle size distribution of semi-autogenous (sag) mill's product. Particulate Science and Technology **36**(7), 867–872 (2018)

[15] Morrell, S., Valery, W.: Influence of feed size on ag/sag mill performance. SAG2001, Vancouver, BC, Canada, 203–214 (2001)

[16] Everett, J., Howard, T., Jupp, K.: Pre-crusher stockpile modelling to minimise grade variability. In: TOS Forum, vol. 5, pp. 35–41 (2015)

[17] Fuerstenau, M.C., Han, K.N.: Principles of Mineral Processing. SME, ??? (2003)

[18] Otsuki, A., Jang, H.: Prediction of particle size distribution of mill products using artificial neural networks. ChemEngineering **6**(6), 92 (2022)

[19] Ghasemi, Z., Neshat, M., Aldrich, C., Karageorgos, J., Zanin, M., Neumann, F., Chen, L.: A hybrid intelligent framework for maximising sag mill throughput: An integration of expert knowledge, machine learning and evolutionary algorithms for parameter optimisation. arXiv preprint arXiv:2312.10992 (2023)

[20] Saldaña, M., Gálvez, E., Navarra, A., Toro, N., Cisternas, L.A.: Optimization of the sag grinding process using statistical analysis and machine learning: A case study of the chilean copper mining industry. Materials **16**(8), 3220 (2023)

[21] Silva, D.H.C., Alves, V.K., Souza, E.S.: Machine learning for particle size prediction in iron ore grinding process. Peer Review **6**(16), 157–177 (2024)

[22] Guo, H., Yan, M., Zhao, J., Wang, L.: Prediction method of particle size qualification rate of high-pressure roller mill based on end-edge-cloud synergy. PeerJ Computer Science **10**, 2151 (2024)

[23] Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S.: Pm2. 5 prediction based on random forest, xgboost, and deep learning using multisource remote sensing data. Atmosphere **10**(7), 373 (2019)

[24] Luo, G.: Online detection system for ore particle size distribution based on deep learning. Journal of Computer, Signal, and System Research **2**(3), 1–16 (2025)

[25] Borgonovo, E.: Sensitivity Analysis: An Introduction for the Management Scientist vol. 251. Springer, Cham (2017)

[26] Ross, S.M.: Introductory Statistics. Academic Press, Amsterdam (2017)

[27] Watson, J., Fitzallen, N., Fielding-Wells, J., Madden, S.: The practice of statistics. In: International Handbook of Research in Statistics Education, pp. 105–137. Springer, Cham (2017)