Powder Technology 420 (2023) 118416

Contents lists available at ScienceDirect

Powder Technology

journal homepage: www.journals.elsevier.com/powder-technology

# Modeling industrial hydrocyclone operational variables by SHAP-CatBoost - A “conscious lab” approach

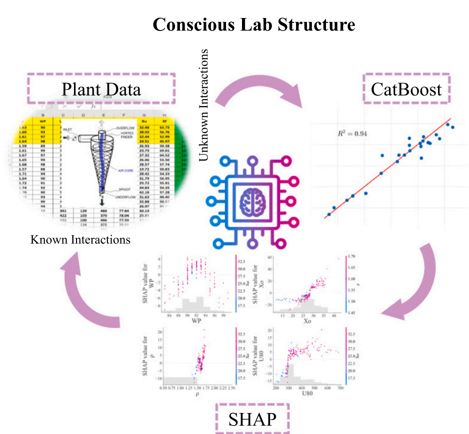
S. Chehreh Chelgani [a,*], H. Nasiri [b,*], A. Tohry [c], H.R. Heidari [d]

[a] Minerals and Metallurgical Engineering, Dept. of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Swedish School of Mines, SE-971 87 Luleå, Sweden
[b] Department of Computer Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran
[c] Research and development unit, Rahbar Farayand Arya Company (RFACo), Tehran, Iran
[d] Deputy of the operation and production, Rahbar Farayand Arya Company (RFACo), Tehran, Iran

## HIGHLIGHTS

- The conscious lab “*CL*” concept was used to model an industrial hydrocyclone.
- SHAP and CatBoost, as novel explainable artificial intelligence, were used for constructing the *CL*.
- Based on SHAP value, overflow solidity had the highest effectiveness on $O_{80}$.
- CatBoost could predict overflow outcomes better than conventional *AI*s.

## GRAPHICAL ABSTRACT




## ARTICLE INFO

*Keywords:*
Hydrocyclone
Ultrafine particles
Random forest
Support vector regression
XGBoost

## ABSTRACT

Undoubtedly hydrocyclones play a critical role in powder technology, which can considerably affect the plants' process efficiency. However, hydrocyclones were rarely modeled on an industrial scale, where a model can be used to train operators and minimize potential scale-up errors and lab costs. The novel approach for filling such a gap would be using conscious lab “*CL*” as a new concept that builds based on an industrial dataset and explainable artificial intelligence (*XAI*). As a novel approach, this study developed a *CL* and explored the interactions between hydrocyclone variables by the most recent *XAI* method called “SHapley Additive exPlanations (SHAP)”, and a novel machine-learning model, “CatBoost”. The hydrocyclone output and the particle size of the plant magnetic separator were modeled by SHAP-CatBoost. SHAP could successfully model all the relationships, and CatBoost could predict the $O_{80}$ and $K_{80,}$ where outcomes had a higher accuracy ($R^2 \sim 0.90$) than other conventional *AI*s.

* Corresponding author.
*E-mail address:* saeed.chelgani@ltu.se (S. Chehreh Chelgani).

https://doi.org/10.1016/j.powtec.2023.118416
Received 5 December 2022; Received in revised form 30 January 2023; Accepted 3 March 2023
Available online 7 March 2023

## 1. Introduction

In mineral processing plants, hydrocyclones are the most typical tools for controlling and classifying particle size distribution (PSD) of powders (fine particles mostly in -500 μm) generated by mills [1]. Since fine grinding of low-grade ore is essential to meet the minerals' liberation degree (vital for the efficient downstream separation processes), the hydrocyclone performance can directly affect the efficiency of upgrading units and the final products' quality. In other words, the PSD of ground ores, besides minerals' properties, would dictate the selection of the upgrading method (such as gravity, magnetic, and froth flotation). Therefore, understanding relationships among hydrocyclone operational variables and their effects on the products' particle size would be crucial for any plant [2–4].

Modeling is a key to better controlling and determining these relationships. Thus, several theoretical models have been developed to explore correlations between operational parameters and model hydrocyclone performances [5,6]. Yet, the generated empirical-numerical models were mostly based on the laboratory scale experiment results [7], and they were limited to specific operating scenarios, could not show the level of variables' effectiveness on the hydrocyclone performance, and assessed each parameter unconnectedly [8–11]. Recently some investigations using artificial intelligence (*AI*) models have been conducted to model hydrocyclone performance (Table 1). Since ANN are mostly black box models, they still hold some fundamental challenges of empirical models. They cannot provide multi-correlation assessments within the operational hydrocyclone variables nor indicate the magnitude of relationships. Moreover, they were mostly based on small laboratory datasets, which would not generate inadequate confidence levels for industrial-scale processes. In some cases, no validation step was considered through those investigations that can develop overfitted models (Table 1).

Considering the "conscious lab (*CL*)" concept could potentially fill all these gaps. Exploring relationships in a simultaneous single and multi-variable assessment using machine learning systems (which would reduce ANN limitations) based on industrial datasets called *CL* [17]. A *CL* can develop a comprehensive insight into the sensitivity of process performance to all monitored operational factors. In detail, *CL* is a new theory that suggests using explainable *AI* (*XAI*) to assess relationships within observed variables from industrial units and model their interactions [17,18]. *XAI* can consider linear and nonlinear relationships through a single-multivariable assessment and rank variables based on their effectiveness on the model outputs. That assessment will also determine the magnitude of relationships. Recently, *CL* has been successfully constructed based on different machine learning models with random forest (RF), support vector regression (SVR), and eXtreme Gradient Boosting (XGBoost)) for various industrial units. This work, as a unique approach, is going to develop a *CL* for an industrial hydrocyclone by using Shapley Additive exPlanations (SHAP) as a novel *XAI* structure.

Each generated model needs to be understandable for operators [19,20]. As a recently developed *XAI*, SHAP can unbox any black box model and translate relationships, bringing them to a human basis. Based on each record's local interpretability, SHAP can thoroughly explain the global average and model output and consider single and multi-correlations for linear and nonlinear interactions. Such an illustration facilitates the training of operators by breaking down *AI* complexity to the general level. In this study, SHAP was coupled with the CatBoost model to enhance its functionality. CatBoost is a newly established *AI* model for gradient boosting on decision trees. It has several advantages over other traditional machine learning methods, such as being faster in training, keeping the original structure of the data frame, and being sparse. Catboost uses the "Ordered target statistics" principle, which solves the prediction shift and target leakage problem [21,22]. The core objective of this work is to emphasize the essential need to understand inter-correlations among hydrocyclone operational

**Table 1**
AI models and their conditions applied for modeling various hydrocyclone performances.

| Model | Input Variables | Output | Data Type | Number of Records | Model Results (Testing Step) | Ref. |
|---|---|---|---|---|---|---|
| Feed-forward neural network | Pressure drop at the inlet, Solid percent vortex Apex diameter | Cut size Overflow Underflow | Laboratory | 19 train<br>6 validation<br>5 test | R-square: 0.98 | [11] |
| Feed-forward neural network | Inlet flow rate<br>Overflow flow rate<br>Underflow flow rate<br>Pressure<br>Volumetric solid concentration<br>Solid density<br>Overflow density<br>Angle of discharge | Cut size | Laboratory | 25 train<br>8 validation<br>8 test | R-square ~ 0.81 | [5] |
| Combination of Self-Organizing Map and Neuro-Fuzzy Inference System | Pressure drop<br>Solid percent<br>Size fraction<br>Overflow densities<br>Underflow densities<br>Water-solid ratios | Cut size | Laboratory | 150 train<br>19 test | Not reported | [12] |
| Back propagated neural network | Overflow densities Underflow densities<br>Apex<br>Spigot flowrates | Cut size | Laboratory | 100 train<br>100 test | R-square ~ 0.97 | [13] |
| Conventional neural network | ImageNet data (300 images) | Underflow particle size | Laboratory | 210 train<br>90 test | R-square ~ 0.85 | [14] |
| Feed-forward neural network | Vortex finder diameter<br>Spigot diameter<br>Inlet pressure | Solid recovery | Laboratory | 100 train<br>50 validation<br>50 test | Not reported | [15] |
| Feed-forward neural network | Inlet flowrate<br>Inlet density<br>Spigot opening<br>Vortex height<br>Temperature of slurry | Cut size | Laboratory | ~100 train<br>~100 test | R-square ~ 0.94 | [16] |

valuables and their effectiveness on the process performance by accurate and human basis methods (*XAI*) on an industrial scale. This approach can be implemented in various processes dealing with powder technology. This study is going to construct a *CL* for a hydrocyclone in the Fakoor Sanat iron ore processing plant using the SHAP-CatBoost model as a narrative approach and assess hydrocyclone's performance on the magnetic separator response (Fig. 1). For comparison principles, random forest (RF), support vector regression (SVR), extreme gradient boosting (XGBoost), and Artificial Neural Network (ANN) as conventional statistical *AI* models, were also examined to assess the SHAP-CatBoost capability for modeling hydrocyclone performance.

## 2. Materials and methods

### 2.1. Dataset

A dataset from the Fakoor Sanat iron ore processing plant (Kerman, Iran) based on daily monitoring of a hydrocyclone for over 6 months (22 December 2019 to 20 June 2020) was gathered to explore the interactions between its operational variables and their effects on the final magnetic concentrators' particle size (magnetite separate from silicates and phosphates) (Fig. 2). The dataset consisted of 185 samples. In the plant, the ball mill's product is subjected to a hydrocyclone, and then, the overflow of the hydrocyclone feeds to medium-intensity magnetic separators (MIMS). The hydrocyclone's cluster consists of four hydrocyclones, and in each case, the main body, apex, and vortex diameter are 660, 220, and 110 mm, respectively. The maximum operating pressure is around 1 bar. Daily measurements were taken, and their mean numbers (for each day) were considered. The parameters (Table 2) of particle size and solid wt% are the mean value of different recorded parameters of representative samples during each shift. The laboratory determined them by sampling and physical tests (wet and dry sieving and filtration, respectively). The working pressure and density of hydrocyclone's feed were monitored from the plant's central control room. SHAP and Pearson's correlation assessed the relationships within the hydrocyclone variables, and its outcome "$O_{80}$" and MIMS particle size "$K_{80}$" were modeled with different machine learning methods.

### 2.2. Pearson correlation coefficient

The Pearson correlation coefficient measures the strength of the linear dependence between two random variables [23]. In other words, the Pearson correlation coefficient measures the closely related two variables and how well they can be described using a linear equation. It is a widely used measure in statistics and is defined as the covariance of the two variables divided by the product of their standard deviations. Pearson correlation coefficient can be computed as follows:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{1}$$

where n represents the sample size, $x_i$ and $y_i$ are sample points, and $\overline{x}$ and $\overline{y}$ denote the sample mean.

### 2.3. Artificial intelligence models

#### 2.3.1. Shapley additive explanations

The Shapley Additive exPlanations (SHAP) is a game theory approach for quantifying the marginal effects of individual features and the interaction/joint effects of two features [24]. It has been demonstrated that SHAP is useful in explaining different machine-learning models [25]. SHAP can explain any machine learning model's output by computing each variable's contribution to the output. This is done by SHAP assigning a score to each variable, which indicates how important it was in making the output. This score is called the SHAP value [26,27]. SHAP satisfies the following properties: local accuracy, missingness, and consistency [28]. Local accuracy means the explanation model should match the original model. The missingness property enforces that missing variables in the dataset are attributed no importance. Consistency indicates that the attribution given to a variable will never decrease, even if we modify a model to increase the impact of a variable on the model. SHAP defines the output of a model as follows:

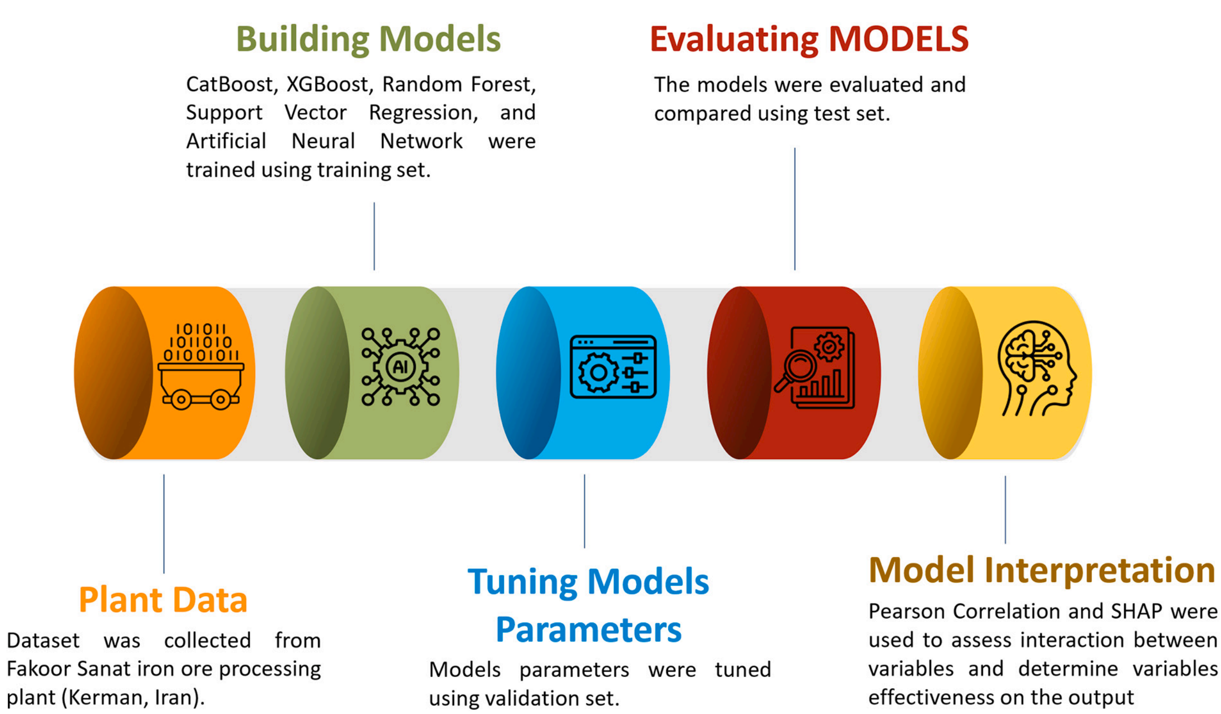



**Fig. 1.** The methodology used for constructing the *CL* for a hydrocyclone using the SHAP-CatBoost model.

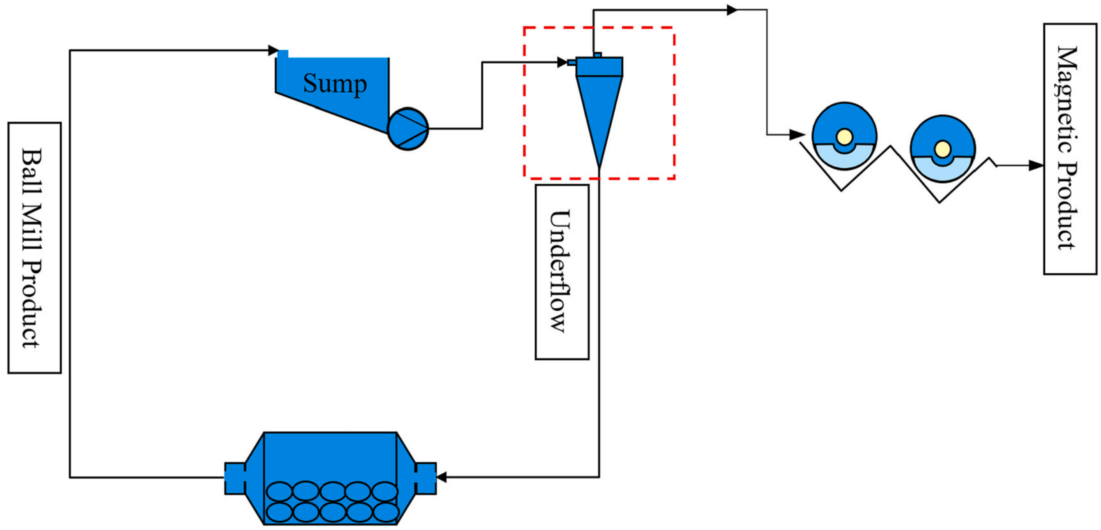



**Fig. 2.** Schematic of the Fakoor Sanat iron ore processing plant.

**Table 2**
Descriptive analyses of the considered hydrocyclone operating parameters.

| Variables | Symbol | Min | Max | Mean | STD |
|---|---|---|---|---|---|
| Feed density (g/cm$^3$) | ρ | 1.35 | 1.74 | 1.59 | 0.08 |
| Feed Solid (wt%) | Xf | 23.87 | 76.11 | 59.86 | 7.63 |
| Working Pressure (kpa) | WP | 85.00 | 99.00 | 91.42 | 2.73 |
| Mill Feed (μm) | $F_{80}$ | 198.15 | 763.72 | 467.98 | 125.50 |
| Underflow (μm) | $U_{80}$ | 203.87 | 684.36 | 357.99 | 98.68 |
| Overflow (μm) | $O_{80}$ | 56.52 | 230.98 | 108.54 | 25.35 |
| Underflow Solid (wt%) | Xu | 37.19 | 80.09 | 74.24 | 4.21 |
| Overflow Solid (wt%) | Xo | 11.99 | 42.18 | 26.48 | 5.67 |
| MIMS products (μm) | $K_{80}$ | 53.05 | 154.40 | 111.44 | 22.07 |

$$f(\mathbf{x}) = g(\mathbf{z}') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i' \quad (2)$$

where $g(\mathbf{z}')$ denotes the SHAP explanation model, $\phi_i$ represents the SHAP value, and $z_i \epsilon \{0,1\}$ is a binary variable. In other words, SHAP decomposes the prediction into a linear function that is comprised of binary variables [29]. Two methods can be used to approximate SHAP values, i.e., Kernel SHAP and Tree SHAP. In this study, Tree SHAP was used to approximate SHAP values as it is much faster than Kernel SHAP and can be used with tree-based methods like CatBoost.

### 2.3.2. CatBoost

CatBoost is a novel variant of the gradient-boosting decision tree method developed by Prokhorenkova et al. in 2018 [21]. Its main idea is that a strong regressor can be created by iteratively combining weak regressors [30]. CatBoost has strong learning capabilities to deal with highly nonlinear data [31,69]. It can be trained by minimizing the expected loss function ($L(.)$) through gradient descent:

$$h^t = \underset{h\epsilon H}{arg\,min} \mathbb{E}\left\{ L\left(y, F^{t-1}(\mathbf{x}) + h(\mathbf{x})\right) \right\} \quad (3)$$

where $y$ denotes the output and $h$ is a gradient step function selected from $H$, i.e., a family of functions. The step function can be calculated as follows:

$$h(\mathbf{x}) = \sum_{j=1}^{J} b_j \mathbb{1}_{\{\mathbf{x} \epsilon R_j\}} \quad (4)$$

where $R_j$ denotes the disjoint regions that correspond to the tree's leaves. $b_j$ is the predictive value of the region, and $\mathbb{1}$ is an indicator function [32].

The expectation in Eq. (3) is approximated by the same dataset, which leads to prediction shift and gradient bias. These issues result in overfitting. CatBoost resolves them by employing the ordered boosting method, which causes improving the robustness and generalization ability of the model [31,32].

### 2.3.3. Extreme gradient boosting

Extreme gradient boosting (XGBoost) is a tree-based ensemble method. It integrates several weak regression trees and sequentially combines their prediction to increase the overall accuracy [33,34]. XGBoost supports parallel and distributed computing and can use the CPU's multithreading feature, which speeds up the computational process [34,35]. Researchers widely use it due to its high robustness and sufficient flexibility [36]. XGBoost adds regularization term to the loss function that smoothens the learned weights and avoids overfitting [37].

XGBoost output can be computed as follows:

$$\hat{y} = \sum_{k=1}^{K} f_k(\mathbf{x}), f_k \,\epsilon\, \Gamma \quad (5)$$

where $\hat{y}$ denotes the predicted output, $\mathbf{x}$ is the input vector, $f_k$ represents the output of the $k$th tree, and $\Gamma$ denotes the function space containing every potential regression tree [29]. Eq. (6) shows XGBoost's objective function:

$$Obj(\theta) = L(\theta) + \Omega(\theta) \quad (6)$$

where $L(\cdot)$ is the loss function that computes the error between the predicted value and target value, and $\Omega(\cdot)$ represents the regularization function that controls the complexity of the model and prevents overfitting [29].

### 2.3.4. Random forest

Random Forest (RF) is a nonparametric, ensemble regression method developed by Breiman in 2001 [38]. RF is a modification of bagging that creates a collection of K-randomized regression trees and averages them. It combines the random subspace method with bagging [39]. The least relevant features can be reduced using RF based on their importance [40].

Solving regression problems by RF has several advantages: low bias, having very few hyperparameters, and minimized risk of overfitting [41]. RF builds an ensemble of K decision trees (DT) as base learners. Each DT predicts the output independently, and then the predictions are averaged to generate the final result (Eq. 7):

$$\widehat{T}(\mathbf{x}) = \frac{1}{K}\sum_{k=1}^{K}\widehat{T}_k(\mathbf{x}) \tag{7}$$

where $\mathbf{x}$ denotes input and $\widehat{T}_k(\mathbf{x})$ is the estimation produced by the $k$th tree [36].

#### 2.3.5. Support vector regression

Support vector regression (SVR), proposed by Drucker [42], is a supervised learning method based on statistical learning theory [43]. SVR is widely used in data-driven problems due to its powerful capability of nonlinear predictions [44]. The advantages of modeling by SVR are: a) It is robust to outliers, b) It has lower computational cost compared to other methods, and c) It is effective when dealing with high-dimensional problems with a small number of samples [45].

The main fundamental concept behind SVR is to map input vectors into an N-dimensional feature space using a mapping function ($\phi(\mathbf{x})$) so that in this space, a linear function can model the relationship between input ($\mathbf{x}$) and output ($y$) as follows:

$$f(\mathbf{x}) = \mathbf{w}\phi(\mathbf{x}) + \mathbf{b} \tag{8}$$

where $\mathbf{w}$ and $\mathbf{b}$ represent weights vector and bias, respectively. Eq. (9) shows SVR's objective function:

$$\phi(\mathbf{w}, \boldsymbol{\xi}) = \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}(\xi_i^- + \xi_i^+) \tag{9}$$

where $C$ represents the regularization constant, and $\xi_i^-$ and $\xi_i^+$ are slack variables, i.e., margin constraints allowed to be violated [46,47].

#### 2.3.6. Artificial neural networks

Artificial Neural Networks (ANN) are mathematical modeling approaches comprised of artificial neurons inspired by research on the human brain system [48,49]. The first computational model of a neuron was proposed by McCulloch and Pitts in 1943 [50]. Due to their good nonlinear mapping capability, ANNs have achieved tremendous success in various domains [51,70]. Formally, a neuron output signal can be computed as follows:

$$O = f(net) = f\left(\sum_{j=1}^{n} w_j x_j + b\right) \tag{10}$$

where $\mathbf{w}$ and $b$ denote weight and bias, respectively, $n$ represents the number of inputs and $f(\cdot)$ is the activation function. Rectified linear unit (ReLU) and Sigmoid are two activation functions that are widely used in ANNs [52,53].

## 3. Results and discussions

### 3.1. Intercorrelations

Pearson correlation "r" was used for the initial single linear intercorrelation assessments within hydrocyclone operational variables throughout the entire dataset (Fig. 3). $F_{80}$ indicated a low positive correlation with $O_{80}$ (r: 0.20), which showed hydrocyclone performance was slightly dependent on the ball mill feed size ($F_{80}$) and mostly relied on the ball mill product size. Pearson assessments also indicated that the working pressure (WP) decreased marginally when bigger particles were fed to the line ($F_{80}$). $F_{80}$ and Xf also showed a positive correlation (r: 0.59). Overflow solid wt% (Xo) demonstrated a substantial positive correlation with $O_{80}$ (r: 0.78).

Pearson correlation (a linear assessment) results showed the potential of several multivariable relationships among the hydrocyclone' variables. The potential of multi-interactions between hydrocyclone operational variables was reported in various investigations [5,12].

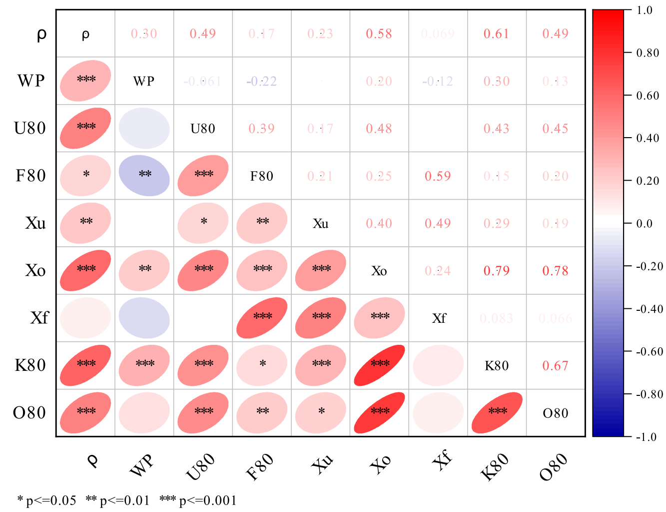



**Fig. 3.** Pearson correlation among hydrocyclone variables.

SHAP assessments can determine these multi-correlations, illustrate them (Fig. 4), rank them based on their importance (Fig. 5), and represent the mean magnitude of their impacts. The SHAP variable importance ranking evaluation revealed a similar pattern to the Pearson correlation assessment (Figs. 3 and 5). SHAP analyses (Fig. 5) demonstrated that Xo has the highest effectiveness on $O_{80}$ modeling. The mean magnitude of this correlation is positive (increasing overflow solid wt% increased the percentage of coarser particles in the overflow "$O_{80}$"). In general, lower solid wt% (lower Xo means higher water content) would transport the fine particles of the slurry to the hydrocyclone center and direct them to the overflow [54].

SHAP also showed the interaction between Xo and WP (Fig. 4). It was reported that increasing WP would be enhanced Xo [55–57]; however, after a certain point, higher WP can partially discharge more water through the overflow and decrease Xo [57]. Farghaly et al. (2020) claimed that by increasing WP, the Xo would decrease, and the cut size could be reduced [58]. SHAP analyses demonstrated only in a specific range (WP: 90–94 kpa and Xo $\geq$ 28%), their interactions had a positive correlation with $O_{80}$ (Fig. 4). Their increase in that range would increase the overflow PSD. Out of these ranges, the SHAP value was negative ($O_{80}$ was finer). It was documented that high WP would increase the turbulence in the hydrocyclone, which can reduce the sharpness of classification by increasing the entrainment of fine particles through coarse particles during their settlement [56,57,59–62].

$U_{80}$ showed the second highest positive effectiveness on $O_{80}$ (Fig. 5). This effectiveness could be explained by its interaction with Xo (Fig. 4). SHAP analyses indicated that for $U_{80}$ below 300 μm and Xo below ~24%, the SHAP value was negative. In contrast, the SHAP value showed a positive effect for their higher values. This relationship can be due to the fact that when the mill product size was generally increased, coarser particles were processed by the hydrocyclone, and subsequently, both $U_{80}$ and $O_{80}$ were raised [63]. It was stated that a fraction of fine particles is regularly subjected to underflow, which supports the negative SHAP value of $U_{80}$ for −300 μm. SHAP mean value assessment also indicated that by increasing ρ, the $O_{80}$ increases (Fig. 5). Although ρ did not show any significant linear correlation (Fig. 3), SHAP indicated its high interactions with other operational variables (Fig. 4) and positive contribution on $O_{80}$ modeling (Fig. 5). Although ρ and Xo has a negative relationship, by increasing ρ and Xo their negative correlations decreased. For ρ and Xo higher than 1.54 (g/cm$^3$) and 26% (respectively), their interactive effectiveness is positive for $O_{80}$ modeling (Fig. 4). It is well understood that unsteady feeding and feed slurry density would be detrimental to the hydrocyclone performances [64]. Increasing density generally causes settling to be hindered, and particle

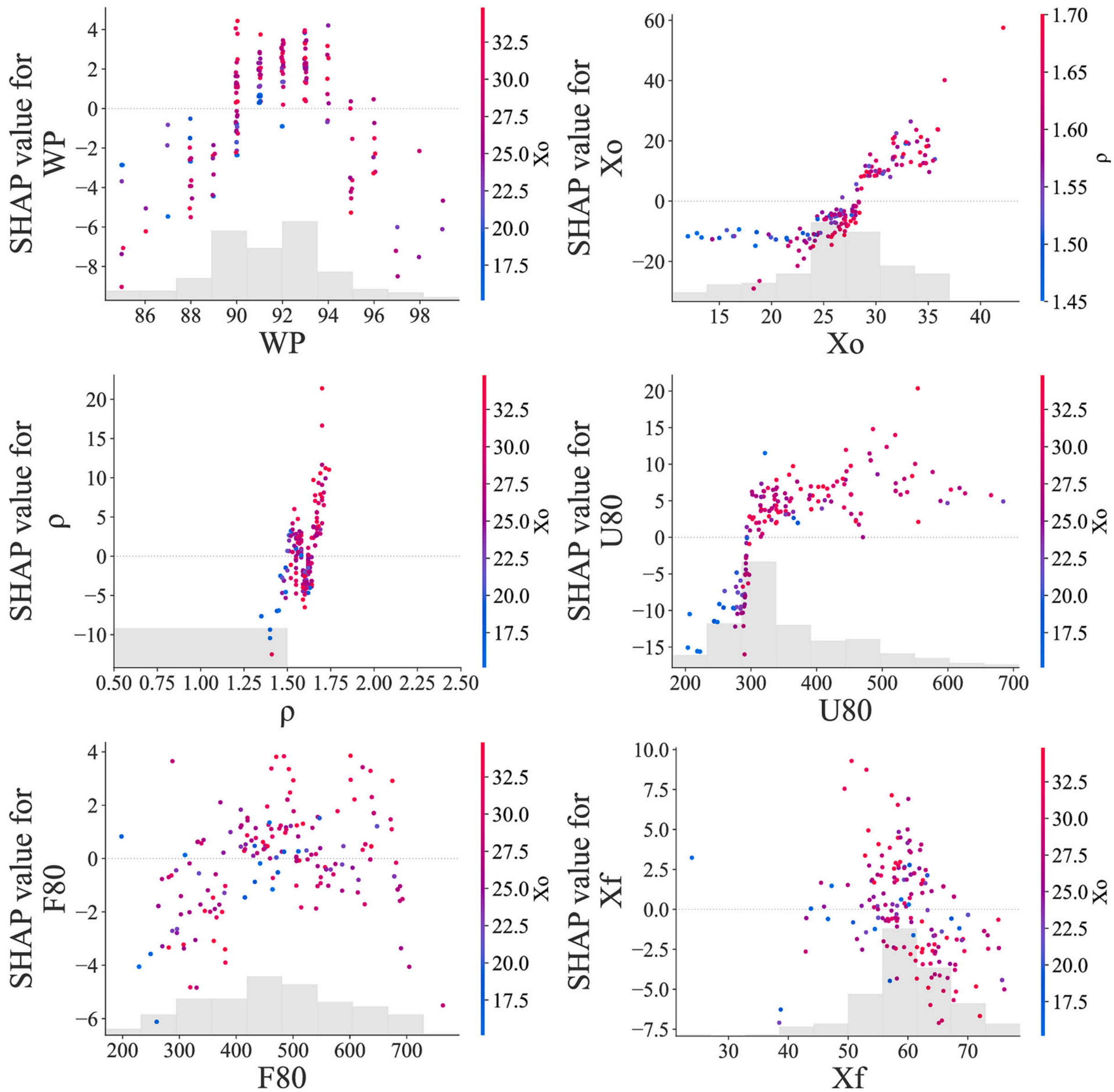



**Fig. 4.** SHAP values for multi-interactions among hydrocyclone operational variables.

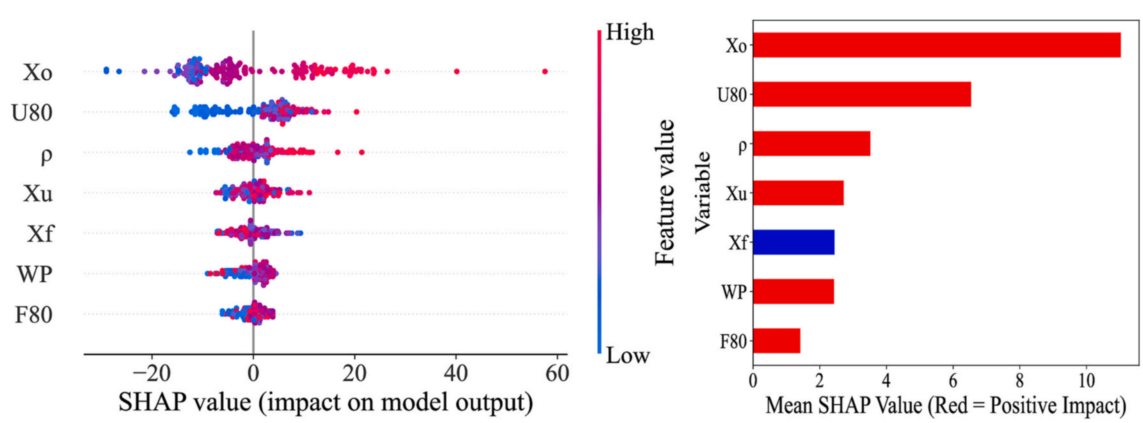



**Fig. 5.** Ranking variables based on their mean SHAP value and their relationship magnitude for $O_{80}$ prediction.

interactions lead to fine particle entrainment [54]. These results indicated that although variables monitored online in the plant's controlling room (WP and ρ) have several interactions with other essential variables, they are not the most effective on hydrocyclone performance. Thus, process operators showed train based on the correlations (linearly and nonlinear) [1] among operational variables and understand interactions (multivariable), something that *CL* has been provided and can have a direct effect on the product properties.

The MIMS operating condition was mostly constant when the process (Fig. 1) was steady in the plant. Thus, its performance could be linked to the hydrocyclone operation and its overflow properties. A meaningful positive relationship (r: 0.67) was observed between $O_{80}$ with $K_{80}$ (Fig. 3). In other words, by increasing the particle size of overflow ($O_{80}$) of the hydrocyclone, the particle size of ($K_{80}$) recovered particles would be increased. Such an interaction would be evident since it meant that coarser particles fed into the MIMS ($O_{80}$) caused an increase in the product particle size ($K_{80}$). Xo had the highest positive linear correlation with $K_{80}$ (r: 0.79). Single and multiple assessments showed a nonlinear multivariable correlation between $O_{80}$ and Xo with $K_{80}$ (Fig. 6). These interactions highlighted the potential of $K_{80}$ prediction based on the hydrocyclone operating parameters, while SHAP ranked the potential effectiveness of all the variables (Fig. 7).

### 3.2. $O_{80}$ and $K_{80}$ prediction

The entire dataset was randomly split into 80, 10, 10% for training, validating, and testing the AI models used to predict the $O_{80}$ and $K_{80}$. Different CatBoost features were examined to find the values of the optimum hyperparameters, adjusted, and tunned the CatBoost model for the training stage. The three main CatBoost hyperparameters are the learning rate, maximum number of trees, and maximum depth of trees. The optimum model parameters were adapted and used for training after the tuning procedure (a try-and-error system) (Table 3). The CatBoost outcomes in the testing stage (Table 4) indicated that this model could accurately estimate the $O_{80}$ and $K_{80}$ based on the plant's operational variables. SHAP could assess all individual records and illustrate each variable contribution level through the prediction (Fig. 8).

Other traditional machine learning models (RF and SVR) and XGBoost were examined to $O_{80}$ and $K_{80}$ with the same training and testing records for comparison purposes. Results (Table 4) released that the CatBoost model can produce the highest accuracy for the $O_{80}$ and $K_{80}$. RF, XGBoost, and CatBoost are ensemble methods. RF is a bagging model, while XGBoost and CatBoost are boosting approaches and require less feature engineering. ANNs can learn complex nonlinear relationships but must adjust several parameters and find an optimal architecture [31]. The kernel function limits SVR. XGBoost suffers from prediction shift and gradient boosting bias, while CatBoost solves prediction shift by employing the ordered boosting method. It also overcomes gradient-boosting biases by using oblivious DTs and applying the same splitting criterion at each tree level. The results obtained by other studies [65–68] also indicated that CatBoost outperformed other ML methods (i.e., SVR, DT, RF, and XGBoost) in the regression task. These results would highlight that SHAP-CatBoost can productively construct a *CL* for a hydrocyclone performance modeling as a commanding *XAI* system. The robust structure of the generated *CL* hypothetically portrayed that this approach could be used for modeling, monitoring, and maintaining different grinding circuits on the industrial scale.

## 4. Conclusion

Constructing models that can show interactions between operational variables on the industrial scale and transforming generated data from complicated knowledge-based systems to human-basis insight would assist plant operators in optimizing production and enhancing process

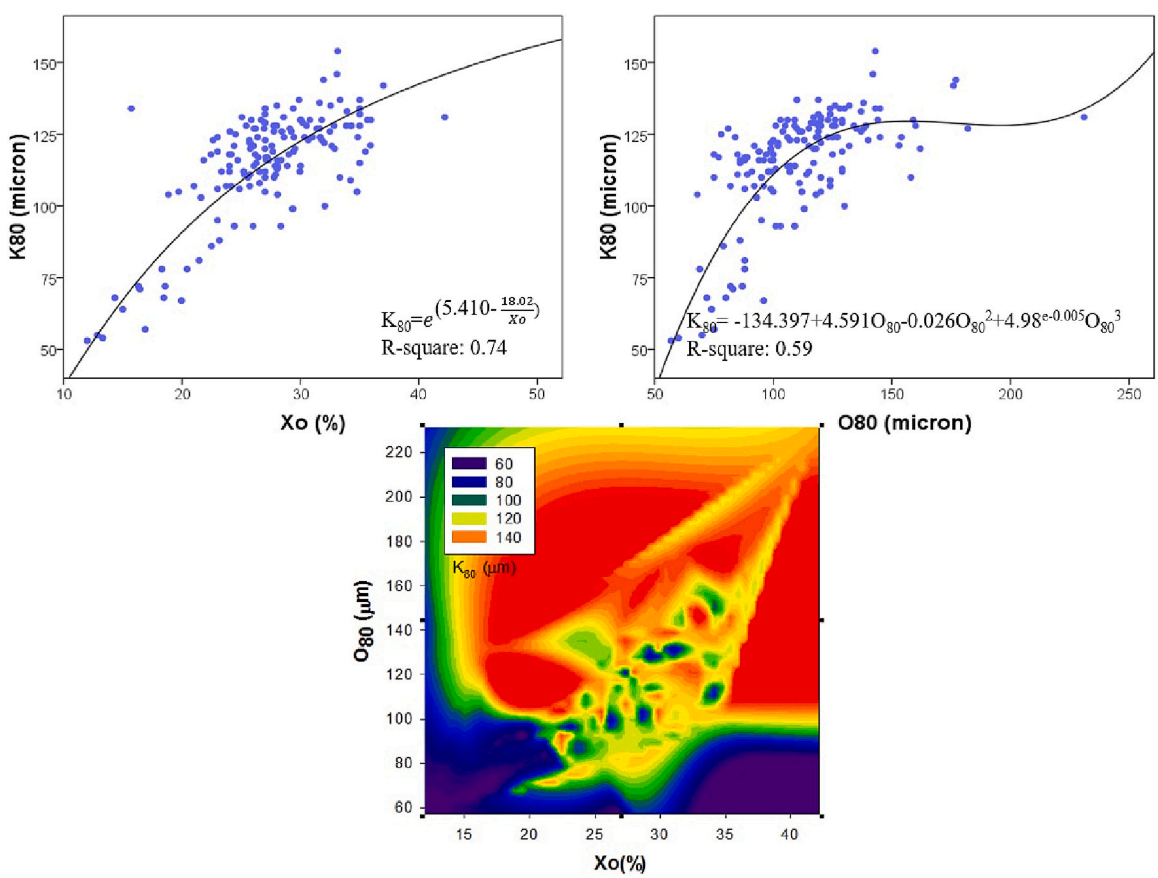



**Fig. 6.** Relationships between overflow variables with MIMS product size.

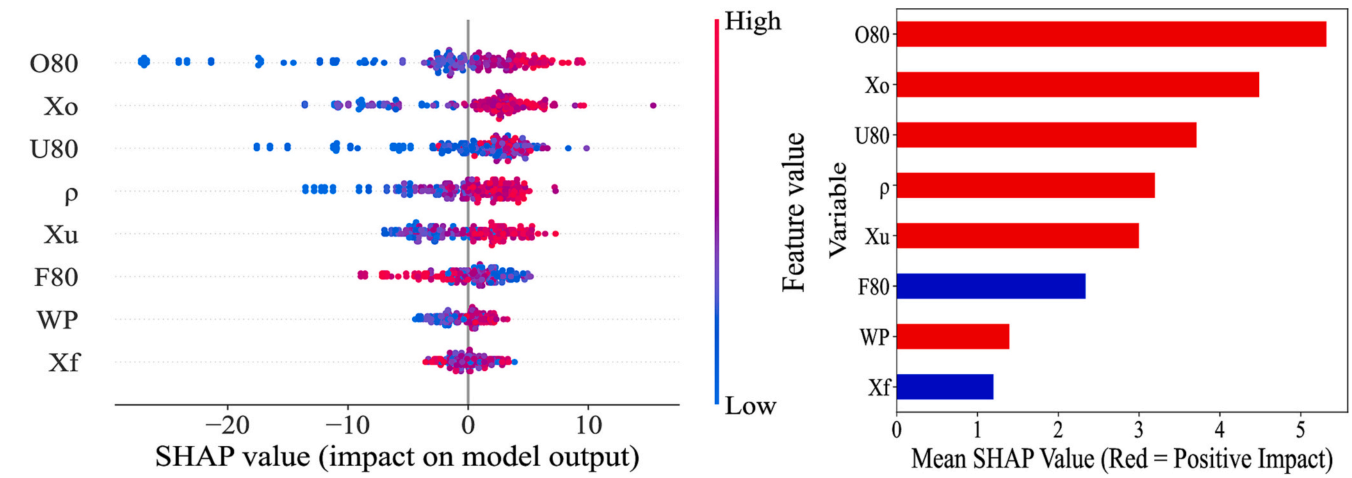



**Fig. 7.** Ranking variables based on their mean SHAP value and their relationship magnitude for $K_{80}$ prediction.

**Table 3**
The parameter settings for the predictive models.

| Parameter | Value ($K_{80}$) | Value ($O_{80}$) |
|---|---|---|
| CatBoost | | |
| Learning rate | 0.897 | 0.394 |
| Maximum number of trees | 105 | 64 |
| Maximum depth of trees | 6 | 6 |
| XGBoost | | |
| Learning rate | 0.285 | 0.279 |
| Maximum number of trees | 26 | 20 |
| Maximum depth of trees | 6 | 6 |
| $\gamma$ | 1.97 | 0.94 |
| Random Forest | | |
| Maximum number of trees | 183 | 8 |
| Support Vector Regression | | |
| Kernel | RBF | RBF |
| Artificial Neural Network | | |
| Activation Funtion | ReLU | ReLU |
| Training Epochs | 1000 | 500 |
| Batch Size | 10 | 10 |

* Radial basis function (RBF) ** Rectified linear unit (ReLU).

efficiency. The outcomes of this investigation indicated that industries that deal with power technology and use hydrocyclones could implement a *conscious lab* for modeling their processes and translate the monitored and recorded dataset to educative information for their operators in the controlling rooms. The generated *conscious lab* in this study indicated several linear and nonlinear multivariable interactions within a hydrocyclone' variables, which affects its performance. Overflow solid wt% has high importance on the overflow product size ($O_{80}$) within all the monitored variables. The magnitude of this effectiveness was positive. It was also indicated that for a steady-state plant, the overflow variables (particle size and solid wt%) of a hydrocyclone can markedly affect the production properties of the downstream magnetic separator ($K_{80}$). Overflow solid wt% and $O_{80}$ showed a significant positive correlation with the magnetic separator product size ($K_{80}$). They demonstrated that their increase would raise the $K_{80}$. The CatBoost model could accurately predict the $O_{80}$ and $K_{80}$ in the testing step ($R^2$: 0.89 and 0.93, respectively). The capability of CatBoost has been argued with typical machine learning methods (random forest and support vector regression) and advanced (XGBoost) *AI* models. The models' outcomes cleared that CatBoost can accurately model the $O_{80}$ and $K_{80}$ (root mean square errors of the models for $O_{80}$ prediction were CatBoost 6.55, XGBoost 6.75, random forest 8.31 and SVR 9.31, and ANN 8.46, and for $K_{80}$ were CatBoost 6.61, XGBoost 7.99, random forest 8.59 and SVR 11.39, and ANN 13.57). In general, these results highlighted that the developed *CL* could be considered for any similar approaches.

**Table 4**
Outcomes of different machine learning methods in various modeling stages for the $O_{80}$ and $K_{80}$ prediction.

| $O_{80}$ | Train | | | Validation | | | Test | | |
|---|---|---|---|---|---|---|---|---|---|
| Method | $R^2$ | RMSE | MSE | $R^2$ | RMSE | MSE | $R^2$ | RMSE | MSE |
| CatBoost | 0.99 | 0.86 | 0.74 | 0.89 | 7.34 | 53.89 | 0.89 | 6.55 | 42.94 |
| XGBoost | 0.98 | 3.51 | 12.35 | 0.89 | 7.73 | 59.74 | 0.88 | 6.75 | 45.51 |
| Random Forest | 0.91 | 7.79 | 60.69 | 0.89 | 7.71 | 59.42 | 0.82 | 8.31 | 68.98 |
| Support Vector Regression | 0.74 | 13.40 | 179.58 | 0.81 | 9.86 | 97.24 | 0.77 | 9.31 | 86.61 |
| Artificial Neural Network | 0.59 | 16.72 | 279.43 | 0.86 | 8.46 | 71.59 | 0.78 | 8.46 | 71.59 |

| $K_{80}$ | Train | | | Validation | | | Test | | |
|---|---|---|---|---|---|---|---|---|---|
| Method | $R^2$ | RMSE | MSE | $R^2$ | RMSE | MSE | $R^2$ | RMSE | MSE |
| CatBoost | 0.99 | 0.0003 | 7.62 | 0.94 | 5.18 | 26.83 | 0.93 | 6.61 | 43.72 |
| XGBoost | 0.99 | 1.0 | 1.09 | 0.92 | 6.35 | 40.33 | 0.90 | 7.99 | 63.97 |
| Random Forest | 0.96 | 4.1 | 16.58 | 0.92 | 6.10 | 37.28 | 0.89 | 8.59 | 73.94 |
| Support Vector Regression | 0.89 | 6.74 | 45.41 | 0.71 | 11.78 | 138.80 | 0.81 | 11.39 | 129.73 |
| Artificial Neural Network | 0.59 | 13.31 | 177.14 | 0.61 | 13.57 | 184.11 | 0.69 | 13.57 | 184.10 |

*Mean Square Error (MSE) **Root Mean Square Error (RMSE).

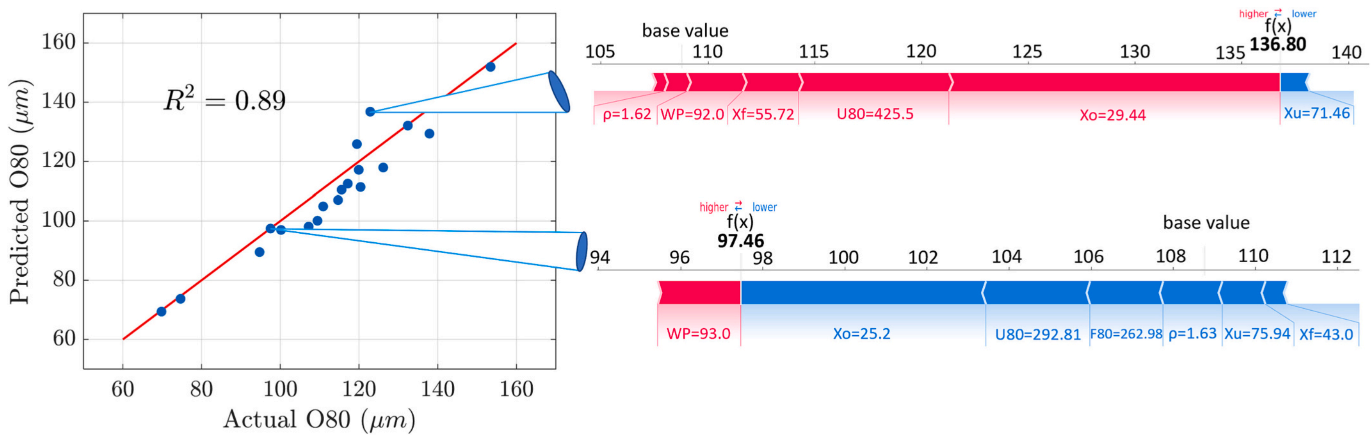



**Fig. 8.** SHAP analyses of each individual record in the test set.

## CRediT authorship contribution statement

**S. Chehreh Chelgani:** Conceptualization, Software, Investigation, Writing – review & editing, Writing – original draft, Methodology, Supervision. **H. Nasiri:** Conceptualization, Investigation, Methodology, Software, Writing - original draft, Writing - review & editing. **A. Tohry:** Conceptualization, Writing – original draft, Methodology, Visualization. **H.R. Heidari:** Supervision, Visualization.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

The data that has been used is confidential.

## Acknowledgment

The authors would like to thank Fakoor Sanat iron ore processing plant (Kerman, Iran) for their support. This manuscript resulted from a project financially supported by CAMM3, the Center of Advanced Mining and Metallurgy, as a center of excellence at the Luleå University of Technology

## References

[1] C. Aldrich, F.D.L. Uahengo, M. Kistner, Estimation of particle size in hydrocyclone underflow streams by use of multivariate image analysis, Miner. Eng. 70 (2015) 14–19.
[2] Z. Zhao, L. Zhou, B. Liu, W. Cao, Computational fluid dynamics and experimental investigation of inlet flow rate effects on separation performance of desanding hydrocyclone, Powder Technol. 402 (2022), 117363.
[3] F. Li, W. Fu, P. Liu, X. Yang, Y. Zhang, X. Li, Y. Zhang, L. Jiang, H. Wang, Investigation of optimal design and separation performance of the hydrocyclone with a vorticose involute-line diversion feeding body, Powder Technol. 117523 (2022).
[4] L. Jiang, P. Liu, Y. Zhang, X. Li, X. Yang, H. Xu, H. Wang, Experimental study of the separation performance of a hydrocyclone with a compound curve cone, Powder Technol. 409 (2022), 117829.
[5] S. Van Loggenberg, G. Van Schoor, K.R. Uren, A.F. der Merwe, Hydrocyclone cut-size estimation using artificial neural networks, IFAC-PapersOnLine. 49 (2016) 996–1001.
[6] D. Bradley, The hydrocyclone, in: The Hydrocyclone, Pergamon, 1965, https://doi.org/10.1016/B978-0-08-010399-0.50001-6.
[7] M.A.Z. Coelho, R.A. Medronho, A model for performance prediction of hydrocyclones, Chem. Eng. J. 84 (2001) 7–14.
[8] M. Yang, R. Jiang, X. Wu, Z. Hu, Y. Yue, Y. Chen, Numerical analysis of flow field and separation characteristics in an oilfield sewage separation device, Adv. Powder Technol. 32 (2021) 771–778.
[9] G. Patra, B. Velpuri, S. Chakraborty, B.C. Meikap, Performance evaluation of a hydrocyclone with a spiral rib for separation of particles, Adv. Powder Technol. 28 (2017) 3222–3232.
[10] D. Hou, Q. Zhao, B. Cui, D. Wei, Z. Song, Y. Feng, Geometrical configuration of hydrocyclone for improving the separation performance, Adv. Powder Technol. 33 (2022), 103419.
[11] M. Karimi, A. Dehghani, A. Nezamalhosseini, S. Talebi, Prediction of hydrocyclone performance using artificial neural networks, J. South. Afr. Inst. Min. Metall. 110 (2010) 207–212.
[12] H. Owladeghaffari, M. Ejtemaei, M. Irannajad, Analysis of hydrocyclone performance based on information granulation theory, ArXiv Prepr. ArXiv0805.2440 (2008).
[13] H. Eren, C.C. Fung, K.W. Wong, A. Gupta, Artificial neural networks in estimation of hydrocyclone parameter d50/sub c/with unusual input variables, IEEE Trans. Instrum. Meas. 46 (1997) 908–912.
[14] J. Olivier, C. Aldrich, Underflow particle size estimation of hydrocyclones by use of transfer learning with convolutional neural networks, IFAC-PapersOnLine. 54 (2021) 85–90.
[15] S. Mohanty, S.K. Das, A.K. Majumder, Artificial neural network modeling and experimental investigation to characterize the dewatering performance of a hydrocyclone, Miner. Process. Ext. Metall. 130 (2021) 396–408.
[16] H. Eren, C.C. Fung, A. Gupta, Application of artificial neural network in estimation of hydrocyclone parameters, in: AusIMM 1996 Annu. Conf, 1996, pp. 225–229.
[17] M. Alidokht, S. Yazdani, E. Hadavandi, S.C. Chelgani, Modeling metallurgical responses of coal Tri-Flo separators by a novel BNN: a "Conscious-Lab" development, Int. J. Coal Sci. Technol. 8 (2021) 1436–1446.
[18] S.C. Chelgani, H. Nasiri, M. Alidokht, Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A "conscious-lab" development, Int. J. Min. Sci. Technol. 31 (2021) 1135–1144, https://doi.org/10.1016/j.ijmst.2021.10.006.
[19] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Adv. Neural Inf. Proces. Syst. 30 (2017) 4765–4774.
[20] M.T. Ribeiro, S. Singh, C. Guestrin, " Why should i trust you?" explaining the predictions of any classifier, in: Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min, 2016, pp. 1135–1144.
[21] L. Prokhorenkova, G. Gusev, A. Vorobev, A.V. Dorogush, A. Gulin, CatBoost: unbiased boosting with categorical features, Adv. Neural Inf. Proces. Syst. 31 (2018).
[22] J. Dutta, S. Roy, OccupancySense: Context-based indoor occupancy detection \& prediction using CatBoost model, Appl. Soft Comput. 119 (2022), 108536.
[23] H. Zhou, Z. Deng, Y. Xia, M. Fu, A new sampling method in particle filter based on Pearson correlation coefficient, Neurocomputing. 216 (2016) 208–215.
[24] X. Wen, Y. Xie, L. Wu, L. Jiang, Quantifying and comparing the effects of key risk factors on various types of roadway segment crashes with LightGBM and SHAP, Accid. Anal. Prev. 159 (2021), 106261.
[25] L. Antwarg, R.M. Miller, B. Shapira, L. Rokach, Explaining anomalies detected by autoencoders using Shapley additive explanations, Expert Syst. Appl. 186 (2021), 115736.
[26] X. Liu, C. Aldrich, Explaining anomalies in coal proximity and coal processing data with Shapley and tree-based models, Fuel. 335 (2023), 126891.
[27] Z. Wen, H. Liu, M. Zhou, C. Liu, C. Zhou, Explainable machine learning rapid approach to evaluate coal ash content based on X-ray fluorescence, Fuel. 332 (2023), 125991.
[28] M.V. García, J.L. Aznarte, Shapley additive explanations for NO2 forecasting, Ecol. Inform. 56 (2020), 101039.
[29] S. Chehreh Chelgani, H. Nasiri, A. Tohry, Modeling of particle sizes for industrial HPGR products by a unique explainable AI tool- a "conscious lab" development, Adv. Powder Technol. 32 (2021) 4141–4148, https://doi.org/10.1016/j.apt.2021.09.020.
[30] Y. Ding, Z. Chen, W. Lu, X. Wang, A CatBoost approach with wavelet decomposition to improve satellite-derived high-resolution PM2. 5 estimates in Beijing-Tianjin-Hebei, Atmos. Environ. 249 (2021), 118212.

[31] K. Deng, X. Zhang, Y. Cheng, Z. Zheng, F. Jiang, W. Liu, J. Peng, A remaining useful life prediction method with long-short term feature processing for aircraft engines, Appl. Soft Comput. 93 (2020), 106344.
[32] N.-H. Nguyen, K.T. Tong, S. Lee, A. Karamanli, T.P. Vo, Prediction compressive strength of cement-based mortar containing metakaolin using explainable categorical gradient boosting model, Eng. Struct. 269 (2022), 114768.
[33] M. Rzychoń, A. Żogała, L. Róg, SHAP-based interpretation of an XGBoost model in the prediction of grindability of coals and their blends, Int. J. Coal Prep. Util. 1–21 (2021).
[34] Y. Feng, L. Liu, J. Shu, A link quality prediction method for wireless sensor networks based on XGBoost, IEEE Access. 7 (2019) 155229–155241.
[35] W. Li, Y. Yin, X. Quan, H. Zhang, Gene expression value prediction based on XGBoost algorithm, Front. Genet. 10 (2019) 1077.
[36] W. Zhang, C. Wu, H. Zhong, Y. Li, L. Wang, Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization, Geosci. Front. 12 (2021) 469–477.
[37] B.S. Bhati, G. Chugh, F. Al-Turjman, N.S. Bhati, An improved ensemble based intrusion detection technique using XGBoost, Trans. Emerg. Telecommun. Technol. 32 (2021), e4076.
[38] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[39] V. Khandelwal, A.K. Chaturvedi, C.P. Gupta, Amazon EC2 spot price prediction using regression random forests, IEEE Trans. Cloud Comput. 8 (2017) 59–72.
[40] J. Abellán-García, J.S. Guzmán-Guzmán, Random forest-based optimization of UHPFRC under ductility requirements for seismic retrofitting applications, Constr. Build. Mater. 285 (2021), 122869.
[41] J. Zhou, E. Li, H. Wei, C. Li, Q. Qiao, D.J. Armaghani, Random forests and cubist algorithms for predicting shear strengths of rockfill materials, Appl. Sci. 9 (2019) 1621.
[42] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V. Vapnik, Others, support vector regression machines, Adv. Neural Inf. Proces. Syst. 9 (1997) 155–161.
[43] Q. Quan, Z. Hao, H. Xifeng, L. Jingchun, Research on water temperature prediction based on improved support vector regression, Neural Comput. & Applic. 1–10 (2020).
[44] R. Tang, C. Fan, F. Zeng, W. Feng, Data-driven model predictive control for power demand management and fast demand response of commercial buildings using support vector regression, Build. Simul. (2022) 317–331.
[45] S.C. Chelgani, B. Shahbazi, E. Hadavandi, Support vector regression modeling of coal flotation based on variable importance measurements by mutual information method, Measurement. 114 (2018) 102–108.
[46] R. Fatahi, H. Nasiri, E. Dadfar, S. Chehreh Chelgani, Modeling of energy consumption factors for an industrial cement vertical roller mill by SHAP-XGBoost: a "conscious lab" approach, Sci. Rep. 12 (2022) 7543, https://doi.org/10.1038/s41598-022-11429-9.
[47] E. Hadavandi, S.C. Chelgani, Estimation of coking indexes based on parental coal properties by variable importance measurement and boosted-support vector regression method, Measurement. 135 (2019) 306–311.
[48] E. Jorjani, A.H. Bagherieh, S. Mesroghli, S.C. Chelgani, Prediction of yttrium, lanthanum, cerium, and neodymium leaching recovery from apatite concentrate using artificial neural networks, J. Univ. Sci. Technol. Beijing, Miner. Metall. Mater. 15 (2008) 367–374.
[49] S.C. Chelgani, S. Makaremi, Explaining the relationship between common coal analyses and afghan coal parameters using statistical modeling methods, Fuel Process. Technol. 110 (2013) 79–85.
[50] A. Xu, H. Chang, Y. Xu, R. Li, X. Li, Y. Zhao, Applying artificial neural networks (ANNs) to solve solid waste-related issues: a critical review, Waste Manag. 124 (2021) 385–402.
[51] H. Nasiri, M.M. Ebadzadeh, MFRFNN: multi-functional recurrent fuzzy neural network for chaotic time series prediction, Neurocomputing. 507 (2022) 292–310, https://doi.org/10.1016/j.neucom.2022.08.032.
[52] S. Ben Jabeur, H. Ballouk, S. Mefteh-Wali, A. Omri, Forecasting the macrolevel determinants of entrepreneurial opportunities using artificial intelligence models, Technol. Forecast. Soc. Change. 175 (2022), 121353.
[53] M.B. Kia, S. Pirasteh, B. Pradhan, A.R. Mahmud, W.N.A. Sulaiman, A. Moradi, An artificial neural network model for flood simulation using GIS: Johor River basin, Malaysia, environ, Earth Sci. 67 (2012) 251–264.
[54] J. Dueck, E. Pikushchak, L. Minkov, M. Farghaly, T. Neesse, Mechanism of hydrocyclone separation with water injection, Miner. Eng. 23 (2010) 289–294.
[55] G. Bundy, How to Fix Common Issues with Hydrocyclones, (n.d.). https://www.townley.net/how-to-fix-common-issues-with-hydrocyclones/ (accessed November 25, 2022).
[56] X. Huang, Y. Lu, G. Wu, Z. Liu, Research on the experiment of the enhancement removal of fine sand by hydrocyclone in sewage treatment plant, Environ. Sci. Pollut. Res. 28 (2021) 337–353.
[57] T. Neesse, J. Dueck, H. Schwemmer, M. Farghaly, Using a high pressure hydrocyclone for solids classification in the submicron range, Miner. Eng. 71 (2015) 85–88.
[58] M.G. Farghaly, H. Attia, H.A. Saleh, A.M. Ramadan, A. Khalek, A combined hydrocyclone-electrocoagulation treatment for different types of industrial wastewater, Physicochem. Probl. Miner. Process. 57 (2021).
[59] J. Dueck, T. Neesse, The hydrocyclone's separation curve in the ultra-fines range, Aufbereitungstechnik. 44 (2003) 17–25.
[60] T. Neesse, J. Dueck, L. Minkov, Separation of finest particles in hydrocyclones, Miner. Eng. 17 (2004) 689–696.
[61] H. Schubert, Which demands should and can meet a separation model for hydrocyclone classification? Int. J. Miner. Process. 96 (2010) 14–26.
[62] T. Neesse, M. Schneider, J. Dueck, V. Golyk, S. Buntenbach, H. Tiefel, Hydrocyclone operation at the transition point rope/spray discharge, Miner. Eng. 17 (2004) 733–737.
[63] J.J. Cilliers, Hydrocyclones for particle size separation, Part. Size Sep. (2000) 1819–1825.
[64] A. Motin, A. Bénard, Design of liquid–liquid separation hydrocyclones using parabolic and hyperbolic swirl chambers for efficiency enhancement, Chem. Eng. Res. Des. 122 (2017) 184–197.
[65] W. Gao, L. Zhou, S. Liu, Y. Guan, H. Gao, J. Hu, Machine learning algorithms for rapid estimation of holocellulose content of poplar clones based on Raman spectroscopy, Carbohydr. Polym. 119635 (2022).
[66] M. Luo, Y. Wang, Y. Xie, L. Zhou, J. Qiao, S. Qiu, Y. Sun, Combination of feature selection and catboost for prediction: the first application to the estimation of aboveground biomass, Forests. 12 (2021) 216.
[67] C. Bentéjac, A. Csörg Ho, G. Mart'inez-Muñoz, A comparative analysis of gradient boosting algorithms, Artif. Intell. Rev. 54 (2021) 1937–1967.
[68] A.A. Ibrahim, R.L. Ridwan, M.M. Muhammed, R.O. Abdulaziz, G.A. Saheed, Comparison of the CatBoost classifier with other machine learning methods, Int. J. Adv. Comput. Sci. Appl. 11 (2020).
[69] M.R. Abbasniya, S.A. Sheikholeslamzadeh, H. Nasiri, Classification of Breast Tumors Based on Histopathology Images Using Deep Features and Ensemble of Gradient Boosting Methods, Comput. Electr. Eng. 103 (2022) 108382, https://doi.org/10.1016/j.compeleceng.2022.108382.
[70] H. Nasiri, et al., Classification of COVID-19 in Chest X-ray Images Using Fusion of Deep Features and LightGBM, in: *2022*, IEEE World AI IoT Congress (AIIoT), 2022, pp. 201–206, https://doi.org/10.1109/AIIoT54504.2022.9817375.