scientific reports

OPEN

Prediction and analysis of dominant factors influencing moisture content during vacuum screening based on machine learning

Ling Nie[1,2], Weiguo Ma[2✉] & Xiangdong Xie[3]

The study of the dominant factors influencing moisture content is essential for investigating vacuum filtration mechanisms. In view of the present situation where there is insufficient experimental data and the dominant factors influencing the moisture content of a filter cake have not been identified, in this study a vacuum filtration apparatus was designed and constructed. Quartz sand particles were used as the filtration material. 300 datasets of moisture contents of a filter cake were obtained under different experimental conditions. Multiple Linear Regression, artificial neural network, decision tree, random forest, and extreme gradient boosting were used to establish a prediction model for moisture content during vacuum screening. By comprehensively analyzing the feature importance rankings and the effects of positive and negative correlations, the dominant factors influencing the moisture content of the filter cake during vacuum screening were the particle ratio, screen mesh, and airflow rate. This finding not only provides a scientific basis for the optimization of vacuum screening technology but also points the way for improving screening efficiency in practical applications. It is of significant importance for deepening the understanding of the vacuum screening mechanism and promoting its extensive application.

The shale shaker is a crucial equipment of the solids control system in oil drilling, tasked with the removal of harmful solid particles from drilling fluid and the recovery of drilling fluid, which plays a significant role in conserving resources and protecting the environment. However, the current high moisture content of cuttings treated by traditional shakers, reaching up to 70%, not only squanders a substantial amount of drilling fluid resources but also exacerbates the environmental impact. In recent years, a novel vacuum screening system based on the principle of vacuum filtration has been proposed[1]. This innovative system can effectively reduce the moisture content of cuttings, enhance separation efficiency, and ensure the full recovery of the liquid phase, thereby achieving a more environmentally friendly treatment process. Consequently, vacuum filtration technology is steadily attracting widespread attention and interest within the industry. Vacuum screening is a solid–liquid separation method based on the principle of filtration. It uses a screen mesh and filter cake as the filtration medium and uses the pressure difference generated by vacuum and air passing through the filtering layer to displace the water in the pores to achieve solid–liquid separation[2]. It is widely used in industries, such as medicine, chemical, and environmental protection, and has a broad engineering application background[3–5]. The moisture content of the filter cake is a critical indicator for evaluating the vacuum filtration performance. Therefore, identifying the dominant factors that influence the filter cake moisture content and establishing a precise prediction model for the moisture content are vital for revealing the vacuum filtration mechanism.

Scholars have investigated the factors influencing the moisture content in vacuum filtration based on the classical filtration theory. Brownell and Katz, Wakeman, Hosten, and Sastry[6–9] established a semiempirical mathematical model of filter cake moisture content based on experimental data and nondimensional capillary numbers. The

[1]School of Computer Science, Yangtze University, Jingzhou 434000, Hubei, China. [2]School of Mechanical Engineering, Yangtze University, Jingzhou 434000, Hubei, China. [3]School of Urban Construction, Yangtze University, Jingzhou 434000, Hubei, China. ✉email: mwg-jh@yangtzeu.edu.cn

results showed that the microstructural parameters of the filter cake influenced the moisture content. Serajuddin et al.[10] performed constant-pressure tests for vacuum filtration of finely ground uranium slurry and derived a power-law equation to predict the solid volume concentration of the filtrate based on experimental data. The results showed that parameters, such as the specific cake resistance and cake permeability, influenced the solid volume concentration in the filtrate. Condie et al.[11] investigated the vacuum filtration process of fine coal slurry using the model proposed by Wakeman. The results showed that the filter cake moisture content was related to the vacuum level, the filter cake thickness and the porosity distribution index. Kerekes and McDonald[12-16] proposed a decreasing permeability model (DPM) under wet-paper-pressing conditions, which showed that the pressure and paper properties influenced the moisture content of the paper. Sjostranda et al.[17] investigated a paper–vacuum filtration process using the DPM numerical model proposed by McDonald and Kerekes. Experimental data from the laboratory machine were fitted to obtain new parameters for the model. A predictive model for the moisture content of the paper during vacuum filtration was developed. The model showed that the paper moisture content was related to the vacuum level, vacuum residence time, and characteristic parameters of the paper (such as initial moisture content, equilibrium moisture content, permeability, and compression factor). However, the establishment of these models requires accurate measurements of the microstructural parameters of the filter cake. This places high requirements for the test instruments and special devices, limiting the practical applications of the results.

Researchers have also studied the factors influencing vacuum filtration using numerical simulation methods. Rezlk et al. established a numerical model of the vacuum filtration process using the level-set method and estimated the moisture content of a paper using the simulation results[18]. The results showed that the moisture content was related to the fiber structure of the paper, the vacuum level, and the vacuum residence time. Li[19] used finite element software to establish a fluid simulation calculation model to simulate the movement of a drilling fluid in a negative-pressure vibrating screen. The study found that the rheological parameters of the drilling fluid and the dynamic parameters of the negative-pressure shale shaker influenced the liquid-phase handling capacity of the shale shaker. Lei[20] analyzed the factors influencing the processing capacity of negative-pressure vibrating screens and found that the processing capacity was related to the drilling fluid viscosity, density, screen mesh, and screen speed. Guo et al.[21] conducted experiments on the filtration and dewatering of gasification slag using a ceramic membrane vacuum filtration system and numerically simulated the dewatering process. Their research results showed that the particle layer moisture content was related to the vacuum level, particle layer thickness, particle equivalent diameter, and the initial moisture content of the particle layer. Ma et al.[22] analyzed the effects of different operating condition parameters on vacuum filtration using a vacuum filtration experiment and FLUENT numerical simulation. The results showed that the vacuum filtration speed was related to the vacuum level, air flow rate, particle size, and filter cake thickness. Numerical simulation methods can be used to analyze the influence of vacuum filtration parameters on the filtration effect; however, the calculation is complex and time-consuming, and the contribution of each parameter cannot be quantified.

In recent years, researchers have investigated the moisture content in vacuum filtration using machine learning algorithms[23,24], such as the support vector machine , artificial neural networks (ANN)[25-27], and multivariate regression. Machine-learning algorithms have advantages over mathematical models based on classical filtration theory and numerical simulation methods, as they can establish predictive models for vacuum filtration in a data-driven manner without relying on physical assumptions and prior knowledge. Guerreiro et al.[28] used a multiple-regression algorithm to predict the moisture content of phosphate concentrate suspensions after filtration. Menezes et al. applied a multiple-regression algorithm to predict the moisture content of particle suspensions during vacuum screening[29]. The results showed that vacuum level, screen inclination angle, and volumetric concentration of solids in the feed affected the moisture content. Huttunen et al.[30] examined the vacuum filtration process of a fluosilicate solution and predicted the filter cake moisture content using standard machine learning algorithms, such as the regularized linear regression algorithms Lasso, Ridge, and ElasticNet, as well as random forest (RF) and gradient boosting. However, the prediction model had low accuracy, and the best-performing algorithm, gradient boosting, predicted the R^2 of the moisture content as approximately 0.8. In addition, the study did not analyze the dominant factors influencing the moisture content. The above studies based on machine-learning algorithms for predicting the moisture content of the filter cake only focused on the prediction performance of the model but paid less attention to the dominant factors influencing the moisture content and the interpretability of the model. In recent years, researchers have coupled machine-learning modules with interpretable Shapley additive explanations (SHAP) modules based on interpretable data-driven models to provide interfaces for reasonable model verification and the quantitative analysis of simulated data[31-34].

Although existing literature has made contributions to modeling the particle layer moisture content during vacuum screening, there are still several limitations. Firstly, traditional mathematical models rely heavily on precise measurements of the microstructure of filter cakes, which not only incurs high costs and strict requirements on experimental equipment, but also limits their practical application due to the scarcity of experimental data. Secondly, while numerical simulations can analyze the influence of vacuum screening parameters on screening efficiency, the computational process is complex and time-consuming, which is unfavorable for rapid prediction and real-time control. Additionally, it is difficult to quantify the specific contributions of each parameter. Furthermore, existing machine learning algorithms exhibit low accuracy in predicting filter cake moisture content and fail to identify the dominant factors influencing particle layer moisture content, while also lacking sufficient interpretability.

In order to address these issues, experiments were conducted using quartz sand particles as the filter material in this study. 300 experimental datasets of the moisture content of the filter cake were obtained under different experimental conditions providing a solid data foundation for establishing a more accurate prediction model. Five machine learning algorithms were employed to establish prediction models for the moisture content of particle layers, and the Grid Search Cross Validation (CV) technique was utilized for hyperparameter tuning,

significantly enhancing the prediction accuracy of the models. Furthermore, Shapley Additive ExPlanations (SHAP) was introduced to enhance the interpretability of the models, and the reasons for the dominant factors were analyzed based on the SHAP values of each feature. The dominant factors influencing the moisture content of the filter cake during vacuum screening were the particle ratio, screen mesh, and airflow rate. The research findings offer invaluable insights and guidance for comprehensively understanding the underlying mechanism of vacuum screening and for enhancing its efficiency and effectiveness.

The innovations and contributions of this study are summarized as follows:

(1) A vacuum screening experimental apparatus was constructed.300 datasets of moisture contents of a filter cake were obtained under different experimental conditions.
(2) To propose a generalized approach that can predict the moisture content during vacuum screening more accurately and efficiently.
(3) The dominant factors influencing the moisture content were identified by combining the feature importance of the ANN, RF, XGBoost, and SHAP models.

## Experimental research
Experimental apparatus
In view of the lack of experimental data of moisture content in vacuum screening, the reliability of the conclusions drawn was poor. Therefore, this study aimed to design and construct a vacuum screening filtration experimental apparatus to collect a large amount of data.

The experimental apparatus was designed and constructed, as shown in Fig. 1. The apparatus comprised a vacuum system and a screening and filtering system. A vacuum pump, vacuum tank, and other connecting and auxiliary equipment were parts of the vacuum system, which provided and maintained a vacuum environment. The screening and filtering system, which filtered and stirred the material, included a material cartridge and a slurry mixing tank.

The tools and instruments used in the experiment included a vortex vacuum pump of the 3RB350-1 type with a maximum airflow rate of 315 m^3/h and a DN600-300-type vacuum tank made of Q235B steel with a total volume of 300 L. Referencing the literature[35], screens with 100–200 meshes are commonly used specifications in engineering. Consequently, three types of API wire-woven screens with meshes of 100, 150, and 200 were selected, each with an effective filtration area of 0.03 m^2. These screens were manufactured by the Xinghuo Metal Mesh Factory in Anping County, China, as shown in Fig. 2. One 304-stainless-steel standard sampling sieve, with meshes of 80, 100, 120 and 200 each, was used, as shown in Fig. 3. One CKLUGD-D50-TD-C vortex flowmeter with a range of 35–380 (m^3/h) was used, that complied with the GB/T 2624-2006 specification, as shown in Fig. 4. The vortex flowmeter, utilizing high-precision sensors for high reliability and powered by a 3.6 V battery, was capable of operating within a temperature range of – 40 to 400 °C. It consisted of a pressure sensor for collecting medium pressure and a flow sensor for measuring the flow rate of the medium, with a gas measurement accuracy of 1.5%. In addition, one Kubei i-2000 digital electronic scale with a range of 0.01–500 g, several beakers, control valves, and measuring cups were used.

Experimental materials
The filter material for vacuum screening should be fine, uniform, and free of impurities, such as quartz sand, inorganic ceramsite, natural zeolite, sludge, soil, etc. Among them, quartz sand was a natural mineral that has low cost, wide availability, and excellent physical and chemical properties. Moreover, quartz sand did not interact
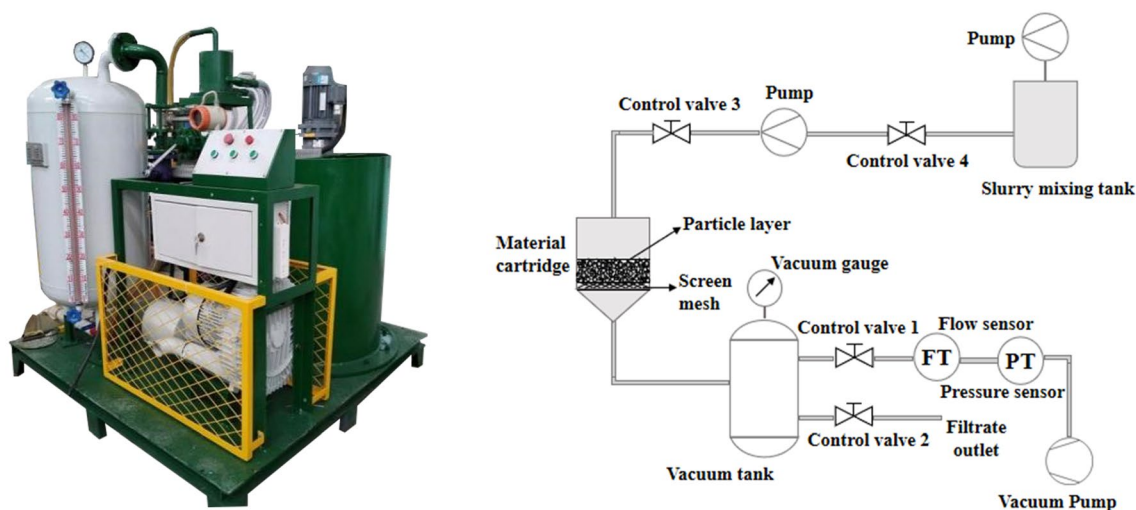
(a) Diagram of the experimental apparatus   (b) Schematic of experimental apparatus

Figure 1. Experimental apparatus for vacuum screening filtration.

Figure 2. Screen.

Figure 3. Sampling sieves.

with the filtrate, provided a consistent filtering effect, and ensured the accuracy of the experimental results. Therefore, quartz sand particles were selected as the filter material. The packed density of the quartz sand used in this study was 1800 kg/m³. As reported in the literature[36], the primary task of traditional vibrating screens is to remove solid particles with diameters greater than 74 μm from drilling fluids. To mimic the different particle size distributions that may be encountered in real-world drilling fluids, the quartz sand particles utilized in this study were carefully selected to fall within a range of 75–180 μm. The particular distribution of particle sizes was meticulously crafted as a blend of various sizes, proportioned relative to the screen mesh opening. After screening using 304-stainless-steel sampling sieves, three types of quartz sand samples with different particle size ranges were selected, as shown in Fig. 5.

Experimental method

The literature[10,11,17–21] reported that various factors influenced the moisture content during vacuum screening, such as airflow rate, vacuum level, particle ratio, particle layer thickness, screen mesh ,and vacuum residence time. In order to investigate the effects of airflow rate, vacuum level, particle ratio, particle layer thickness, and screen mesh on the moisture content of the vacuum screening, this experiment was conducted under standard-ized conditions, specifically at an ambient indoor temperature of 23 °C. Three types of quartz sand samples were mixed to different proportions to form a mixed sample, as listed in Table 1. Appropriate amounts of water were
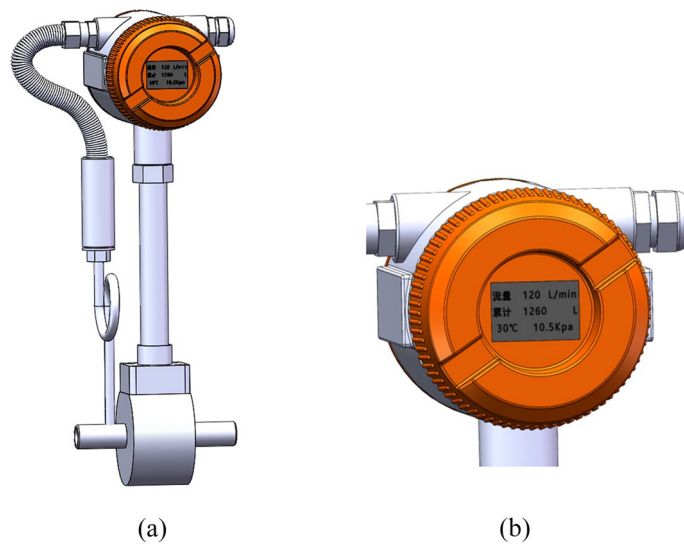
Figure 4. (**a**) vortex flow meter; (**b**) vortex flow meter gauge.

Figure 5. Quartz sand.

Particle ratio	Particle diameter		
	75–120 μm (%)	120–150 μm (%)	150–180 μm (%)
Ratio 1	10	40	50
Ratio 2	20	30	50
Ratio 3	30	20	50

Table 1. Particle size mix ratio.

then added to fully soak these samples, resulting in wet particle samples. Next, 200–600 g of wet particles of different masses were weighed and uniformly distributed across the screen, and the thickness of the particle layer was accurately calculated using Eq. (1), forming the experimental parameters listed in Table 2.

$$L = \frac{m}{\rho \cdot s}$$

(1)

where ρ represents the bulk density of particles, s represents the surface area of the particle layer, and m represents the mass of the particle layer.

Particle mass (g)	Particle layer thickness (mm)	Particle ratio	Screen mesh	Residence time (s)	Number of experiments
200	3.5	1, 2, 3	100, 150, 200	5, 10, 15	3
300	5.3	1, 2, 3	100, 150, 200	5, 10, 15	3
400	7.1	1, 2, 3	100, 150, 200	5, 10, 15	3
500	8.8	1, 2, 3	100, 150, 200	5, 10, 15	3
600	10.6	1, 2, 3	100, 150, 200	5, 10, 15	3

Table 2. Experimental parameters.

The experiments focused on particle layers with varying masses and particle ratios. The control valve1 in Fig. 1b was used to adjust the vacuum level and airflow rate for each experiment. Each group of experiments employed a unique screen mesh. As the particle mass changed, so did the thickness of the layer. The moisture content of the filter cake was determined to be dependent on the specific experimental conditions used.

The moisture content of the filter cake was determined following the following procedure, as shown in Fig. 6.

(1)   Weigh the empty volumetric flask: First, The mass of an empty 50 mL volumetric flask was weighed, denoted m_0.
(2)   Add dry particles and Weigh: Next, 50 mL of dry particles (quartz sand) was added to an empty volumetric flask, and its mass was denoted m_1.
(3)   Prepare wet particles for initial weighing:Subsequently, 50 mL of wet particles was randomly added to an empty volumetric flask before filtration, and its mass was denoted m_2.
(4)   Distribute wet particles on screen: A specific mass of wet particles was distributed to the screen of the material cartridge, and the height of the wet particle layer was denoted L.
(5)   Vacuum filtration setup: The vacuum pump was turned on, and the vacuum in the vacuum tank was adjusted until it reached the set value.
(6)   Filtration and data collection: The timer was started, and data were recorded at regular intervals. When the design experiment time was reached, 50 mL of wet particles were randomly extracted after filtration; its mass was denoted m_3.The airflow rate and pressure gauge value of the vacuum tank were recorded after the vacuum level reached the set value and stabilized for 5 s. Subsequently, 50 mL of wet particles after filtration were randomly sampled and weighed. The data were recorded, and the average value was calculated for each group of experiments after repeating the experiment three times under the same conditions.

The experiments were conducted on samples with different particle size ratios using the same method. The initial and final moisture contents after filtration were obtained using Eqs. (2) and (3), respectively.

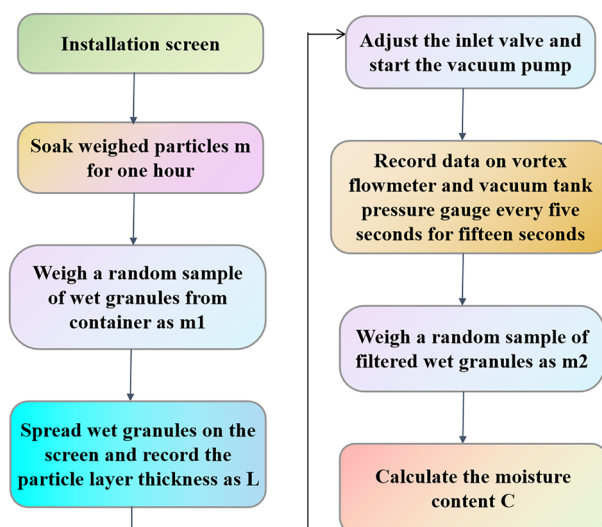$$C_0 = \frac{m_2 - m_1}{m_1 - m_0} \times 100\% \tag{2}$$

Figure 6. Experimental procedure.

$$C = \frac{m_3 - m_1}{m_1 - m_0} \times 100\% \qquad (3)$$

Discussion of experimental results

Based on the experimental method described in Section "Experimental method", each experiment was repeated three times, and the average value was obtained, screening the data and discarding the anomalies, resulting in 300 sets of experimental data. The initial moisture content of the particle layer for each experiment calculated using Eq. (1) was 37.14%, and the moisture content after filtration ranged between 3 and 17%. Analysis of the experimental data revealed that different experimental parameters had varying degrees of influence on the moisture content of the filter cake.

Effect of airflow rate on moisture content

Figure 7 shows the relationship between the moisture content of filter cake and the airflow rate. The experimental parameters were set as follows: 150-screen mesh, particle ratio of 1, and particle layer thickness values of 3.5 and 5.3 mm. With an increase in the vacuum residence time, the moisture content of the filter cake showed a rapid decrease initially, followed by a slow decline (Fig. 7). In addition, as the airflow rate increased, the moisture content decreased. The moisture content curve corresponding to a particle layer with a thickness of 3.5 mm is shown in Fig. 7a. When the vacuum residence time reached 5 s, the moisture content rapidly decreased from 37.14 to approximately 10%. However, during the vacuum residence time of 5–15 s, the moisture content decreased by approximately 2%. This is because, under the action of vacuum airflow at a residence time of 5 s, most of the liquid in the pores of the filter cake was transported by the airflow. As the airflow rate increased, the flow rate of the liquid in the pore spaces of the filter cake also increased, resulting in a decrease in the moisture content of the filter cake. Under the same experimental conditions, increasing the airflow rate decreases the moisture content, indicating that the airflow rate influences the moisture content.

Effect of screen mesh on moisture content

Figure 8 shows the relationship between the moisture content of the filter cake and the screen mesh. The experimental parameters were set as particle ratios of 1 and 2, a particle layer thickness of 5.3 mm, and a vacuum level of 20 kPa. With increasing vacuum residence time, the moisture content initially decreased rapidly and then decreased gradually. As the screen mesh size increased, the moisture content correspondingly increased. For example, when the experimental parameters were set with a particle ratio of 1 and particle layer thickness of 5.3 mm, the corresponding moisture content curve was as shown in Fig. 8a. When the residence time was 15 s, and the screen mesh size was 100, the moisture content was 4.09%. However, when the screen mesh size was increased to 200, the moisture content reached 8.52%. This is because an increase in the screen mesh size leads to an increased resistance of the liquid passing through the screen, thereby decreasing the flow rate of the liquid and increasing the final moisture content of the filter cake. This indicates that under the same experimental conditions, reducing the screen mesh decreases the moisture content, demonstrating that the screen mesh influences the moisture content.

Effect of particle ratio on moisture content

Figure 9 shows the relationship between the moisture content and particle ratio. The experimental parameters were set as follows: particle layer thickness of 5.3 mm, vacuum level of 15 kPa, and screen mesh sizes of 100 and 150. As the residence time increased, the moisture content initially decreased rapidly and then gradually
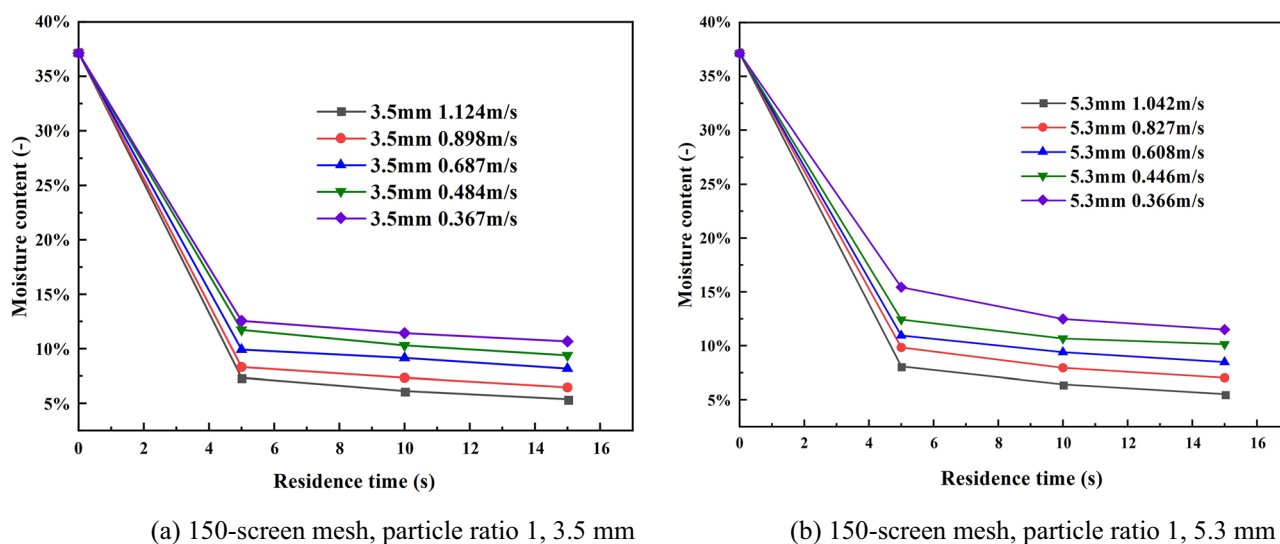

(a) 150-screen mesh, particle ratio 1, 3.5 mm   (b) 150-screen mesh, particle ratio 1, 5.3 mm

Figure 7. Effect of airflow rate on moisture content.

(a) 5.3 mm, 20 kPa, particle ratio 1

(b) 5.3 mm, 20 kPa, particle ratio 2

Figure 8. Effect of screen mesh.
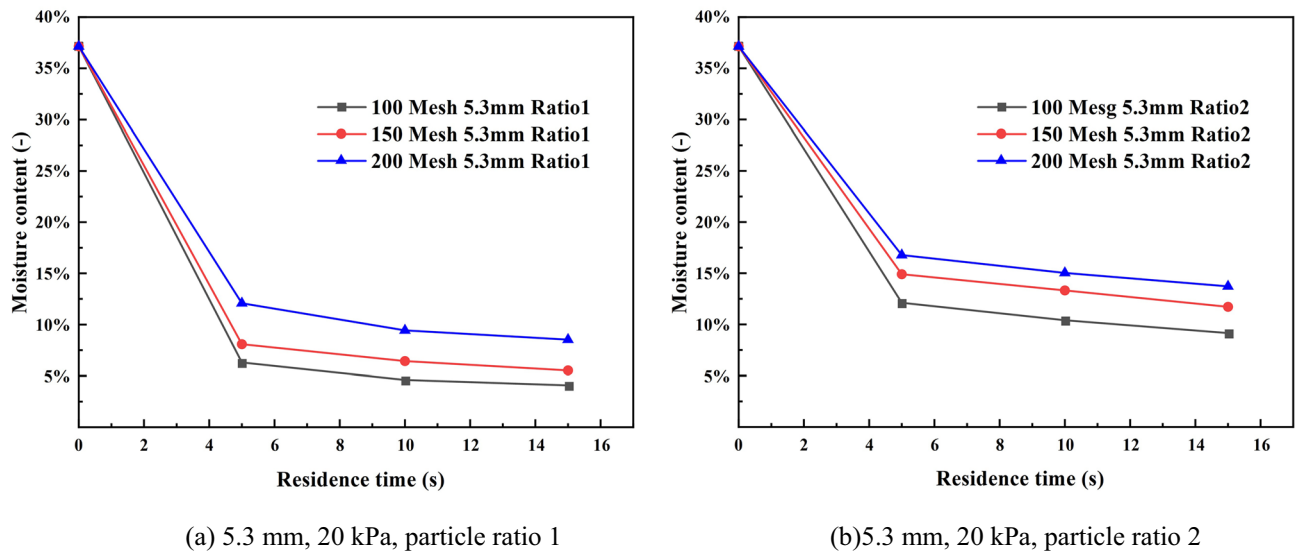

(a) 100 screen mesh, 5.3 mm, 15 kPa

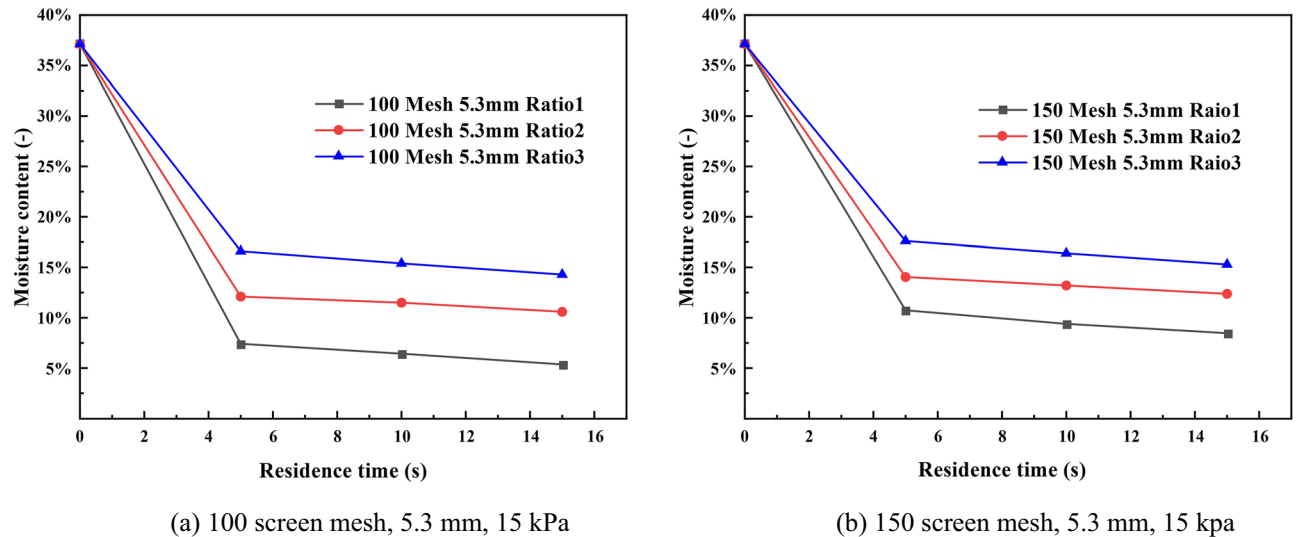(b) 150 screen mesh, 5.3 mm, 15 kpa

Figure 9. Effect of particle ratio.

decreased at a slower rate (Fig. 9). With an increase in the particle ratio, the moisture content increased. For example, when the experimental parameter was set with a screen mesh of 100, the corresponding moisture content curve was as shown in Fig. 9a. At a residence time of 15 s, particle ratios of 1, 2, and 3 resulted in moisture contents of 5.37%, 10.59%, and 14.29%, respectively. The particle layer with a ratio of 1 had approximately 9% lower moisture content than that with a ratio of 3. This is because, as the particle ratio increased, the resistance of the liquid passing through the pores in the particle layer increased, leading to a decrease in the liquid flow rate and an increase in the final moisture content. This indicates that under the same experimental conditions, decreasing the particle ratio decreases the moisture content of the filter cake, and the particle ratio significantly influences the moisture content.

Effect of particle layer thickness on moisture content
Figure 10 shows the relationship between the moisture content and particle layer thickness. The experimental parameters were set as follows: vacuum level of 20 kPa, screen mesh of 100, and particle ratios of 1 and 2. As the residence time increased, the moisture content initially decreased rapidly and then gradually decreased at a slower rate. In addition, the moisture content increased correspondingly with increasing particle layer thickness. For example, when the residence time was 15 s, the moisture content with a thickness of 3.5 mm was approximately 4% lower than that with a thickness of 10.6 mm (Fig. 10b). This is because, as the particle layer thickness increases, the resistance of the liquid passing through the pores in the particle layer also increased, leading to a decrease in the liquid flow rate and an increase in the final moisture content. This indicates that under the same
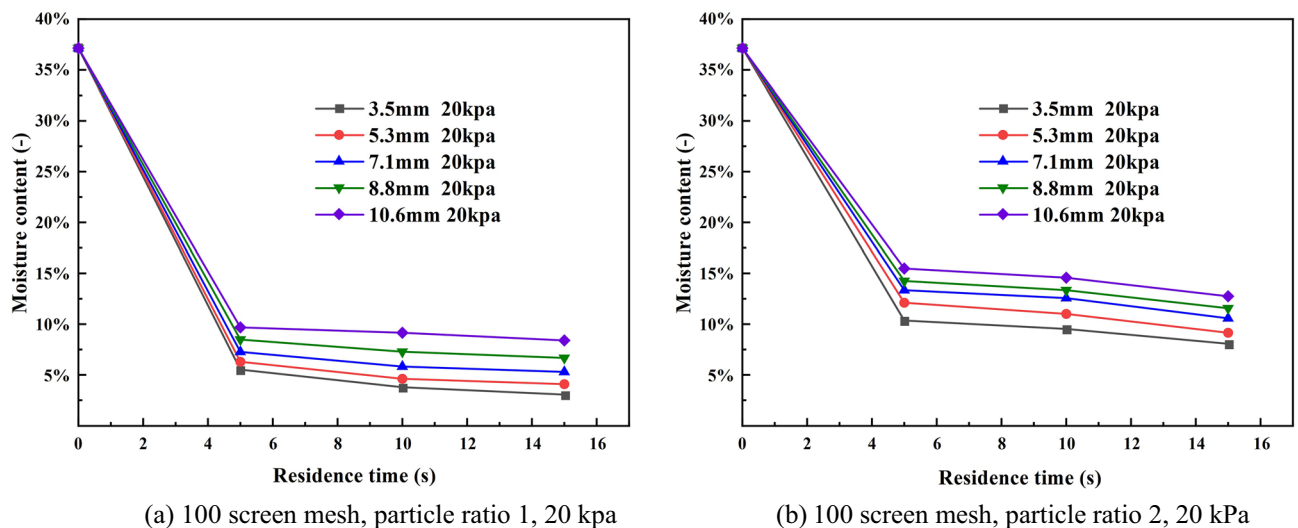
(a) 100 screen mesh, particle ratio 1, 20 kpa

(b) 100 screen mesh, particle ratio 2, 20 kPa

Figure 10. Effect of particle layer thickness.

experimental conditions, decreasing the particle layer thickness decreases the moisture content, highlighting the effect of the particle layer thickness on the moisture content.

The experimental results revealed that the filter cake moisture content during vacuum screening was influenced by various factors, such as the airflow rate, screen mesh, particle ratio, particle layer thickness, and residence time. However, the specific effects of these factors on moisture content, their interplay, and how they jointly determine the final moisture content during vacuum screening may be difficult to fully and accurately quantify in experimental analysis.

During the process of vacuum screening, the relationships between various factors and moisture content are often complex and non-linear. Traditional statistical methods may find it difficult to accurately capture these non-linear relationships, while machine learning models have powerful data processing and pattern recognition capabilities. By training machine learning models, complex nonlinear relationships can be automatically extracted from experimental data to reveal hidden patterns and regularities, thereby improving the accuracy of moisture content prediction. This is of great significance for real-time monitoring and rapid adjustment during the vacuum screening process, which helps to improve product quality and production efficiency.

Additionally, the feature importance method in machine learning can quantify the contribution of each factor to the moisture content. This quantitative analysis not only helps to understand the intrinsic mechanism of the vacuum screening process more deeply, but also provides a more scientific basis for the optimization of the vacuum screening process. Therefore, machine learning methods would be used to predict the moisture content during vacuum screening, and the feature importance of the machine learning model would be used to analyze the dominant factors influencing the particle layer moisture content during vacuum screening.

Prediction of moisture content for vacuum screening

In the present study, five different machine learning algorithms, namely: multiple linear regression(MLR), decision tree (DT), artificial Neural Network (ANN)[37], random forest (RF)[38], and eXtreme Gradient Boosting (XGBoost)[39] were built and compared against each other in terms of predicting moisture content during vacuum screening.

The experimentally determined particle ratios, particle layer thickness, screen mesh, airflow rate, vacuum level, and residence time were used as inputs to accurately predict the moisture content of the filter cake during vacuum screening, and the moisture content was used as the output. The evaluation metrics for the model performance included the mean absolute error (MAE), root-mean-square error (RMSE), and coefficient of determination (R^2). The objective was to determine an optimal model that resolves the dynamic prediction of moisture content in vacuum screening. An machine-learning experiment was conducted on the Anaconda platform using the Python 3.8 programming language, and NumPy, Pandas, and Scikit-learn packages were called to implement machine-learning algorithm modeling[40]. The experimental data were randomly divided into training and testing sets in a ratio of 8:2.

Machine learning algorithms

Five machine learning algorithms were utilized to predict the moisture content during vacuum screening, namely: multiple linear regression(MLR), decision tree (DT), artificial Neural Networks (ANN), random forest (RF), and eXtreme Gradient Boosting (XGBoost). Thus, the theoretical interpretation of these algorithms was illustrated in the following subsections.

Multiple linear regression

Multiple Linear Regression (MLR) is a commonly used statistical method for analyzing the linear relationship between two or more independent variables and a dependent variable. The MLR model is simple and easy to understand, with high computational efficiency, making it suitable for modeling linear relationships. As a benchmark model, MLR can help us initially understand the linear relationships within the data and provide a reference for other more complex models.

Decision tree

Decision Tree (DT) is a machine learning algorithm that mimics the human decision-making process. It recursively partition the data into smaller subsets based on a series of feature-based binary questions, continuing until it reaches leaf nodes that can make predictions. Its advantages lie in the strong interpretability of the model, which can intuitively demonstrate the relationship between data features and target variables. In addition, decision trees can handle data with nonlinear relationships among features and also show good adaptability to small sample sizes.

Artificial neural networks

Artificial neural networks (ANN) are algorithms designed to mimic the way the human brain receives and processes information. ANN do not require a predetermined mathematical equation defining the mapping relationship between inputs and outputs. The advantage of ANN lies in their ability to capture complex nonlinear relationships, with strong learning capabilities and adaptability, which are widely applied to complex pattern recognition and predictive problems[41]. Since the moisture content during vacuum screening was influenced by multiple factors, which might be nonlinear relationships between these factors. Therefore, this study employed ANN to capture these complex nonlinear relationships. Artificial neural networks were used to establish a prediction model for moisture content during vacuum screening. The specific structure of the model was shown in Fig. 11.

Random forest

Random Forest is an ensemble learning algorithm that uses decision trees as base learners. Following the RF bagging method reduces the chances of results being affected by outliers. In this algorithm, each decision tree is trained based on a randomly selected subset of samples and features. Then, the prediction results of multiple decision trees are integrated through voting or averaging to obtain the final prediction result. Its advantage lies in the integration of multiple decision trees, which endows it with strong generalization ability and resistance to overfitting. Compared to some other 'black box' machine learning models, the Random Forest algorithm offers better model interpretability, enabling it to identify the features that have the greatest impact on the prediction results, thereby helping us understand the dominant factors that influence the moisture content during vacuum screening.

XGBoost

XGBoost, standing for Extreme Gradient Boosting, is an optimized implementation of Gradient Boosting designed as as a regularized, faster, and more accurate version to handle the overfitting problem. XGBoost utilizes parallel computing to optimize the training process. Contrary to Random Forest, where trees are grown simultaneously, XGBoost constructs trees in a sequential manner, with each tree being built consecutively following
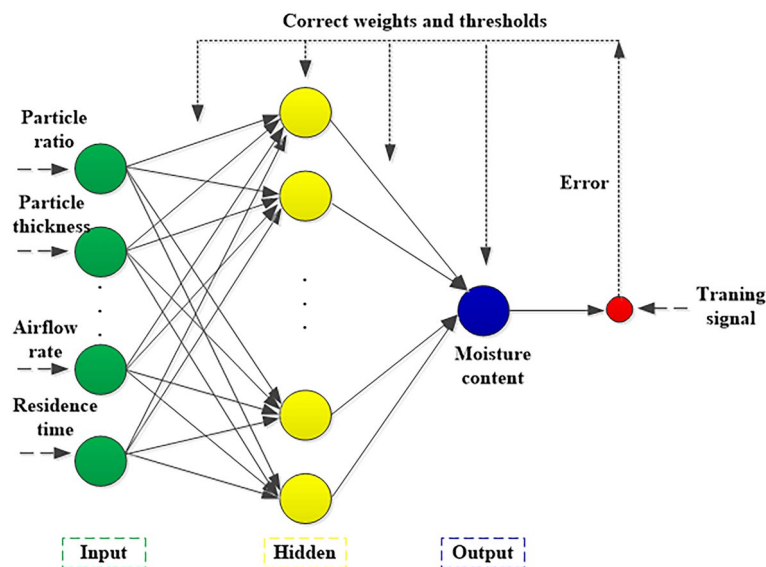

Figure 11. Three-layer ANN neural network model.

the previous one. This approach ensures that the overall training process is more efficient and effective. XGBoost is known for its high predictive accuracy and robustness, capable of handling complex nonlinear relationships. It demonstrates remarkable performance when handling large-scale datasets and complex features[40], thereby making it an ideal choice for fulfilling the high-precision predictive requirements of this study.

Model parameter tuning

The hyperparameters were optimized using the fivefold cross-validation method to improve the performance of the machine-learning models. In addition, the performance of the models was evaluated to select the optimal hyperparameters. The range of hyperparameters for various machine learning models and their optimal values after grid search cross-validation were presented in Table 3. Figure 12 displayed the validation curves for the adjustment of the critical hyperparameters of the four models.

As for ANN, the number of hidden layer nodes is a crucial hyperparameter[42,43], as it directly influences the network's capacity and ability to learn complex patterns. Determining the appropriate number of hidden layer nodes is essential for balancing the network's generalization ability and avoiding overfitting, while also affecting training efficiency and the final performance of the model.As clearly demonstrated in Fig. 12a, upon setting the number of hidden layer nodes to 9, the coefficient of determination (R^2) for both the training and validation sets attains a peak and subsequently plateaus, conclusively suggesting that 9 represents the optimal number of hidden layer nodes.

As for the Decision Tree, Random Forest, and XGBoost models, "max_depth" is a critically important hyperparameter, which controls the maximum depth of the trees[40,44]. It plays a pivotal role in curbing overfitting, thereby ensuring that the model maintains a balance between complexity and generalizability. Additionally, "max_depth" significantly influences the model's training velocity and its ultimate predictive precision.

Model prediction and comparative analysis

The performances of five machine learning models with optimized hyperparameters were compared and analyzed to select the best moisture-content prediction model using the RMSE, MAE, and R^2. Table 4 presents a comparison of the results of the different models for each indicator. The data showed that the XGBoost model achieved R^2, RMSE, and MAE values of 0.9605, 0.0058, and 0.0041, respectively, for the test set (Table 4). Compared to the MLR prediction model, the R^2 value increased by 34.56%, the RMSE decreased by 56.72%, and the MAE decreased by 69.17%. The XGBoost model also outperformed the RF model, which is another ensemble tree type, by 3.67% (R^2), 26.58% (RMSE), and 36.92% (MAE).

The prediction results of the three models on the training and test datasets can be observed intuitively in Fig. 13. Except for a considerable deviation shown by the MLR model in Fig. 13a on both the training and testing datasets, the other four models exhibit a close match between their predicted and experimental values. In particular, the XGBoost model achieved an R^2 value of 99.36% on the training set and 96.05% on the testing set. The XGBoost model outperformed ANN and RF in terms of prediction and generalization ability. Overall, the XGBoost model demonstrated a significant advantage over the MLR, ANN, DT and RF models in terms of performance.

Model	Hyperparameter	Selected range	Optimal value
ANN	Input variables	6	6
	Number of hidden layer	[1–5]	1
	Hidden layers nodes	[3–12]	9
	learning_rate	[0.001, 0.002, 0.003, 0.004]	0.001
	Epoch	[100, 200, 300, 400, 500, 600, 700, 800, 900, 1000]	200
	Activation funcon	tanh, ReLU, trainlm	tanh
Decision Tree	min_samples_leaf	[5–10]	10
	max_depth	[1–10]	5
	Max _features	[2–6]	4
Random Forest	n_estimators	[100, 120, 140, 160, 180]	140
	max_depth	[1–10]	4
	Max _features	[2–6]	3
XGBoost	n_estimators	[100, 200, 300, 400, 500]	300
	max_depth	[1–10]	5
	learning_rate	[0.1, 0.2, 0.3, 0.4, 0.5]	0.3
	Gamma	[0, 1, 2, 3, 4, 5]	0
	Minimum child weight	[1–5]	2
	reg_lambda	[0.001, 0.0011, 0.0012, 0.0013]	0.0012

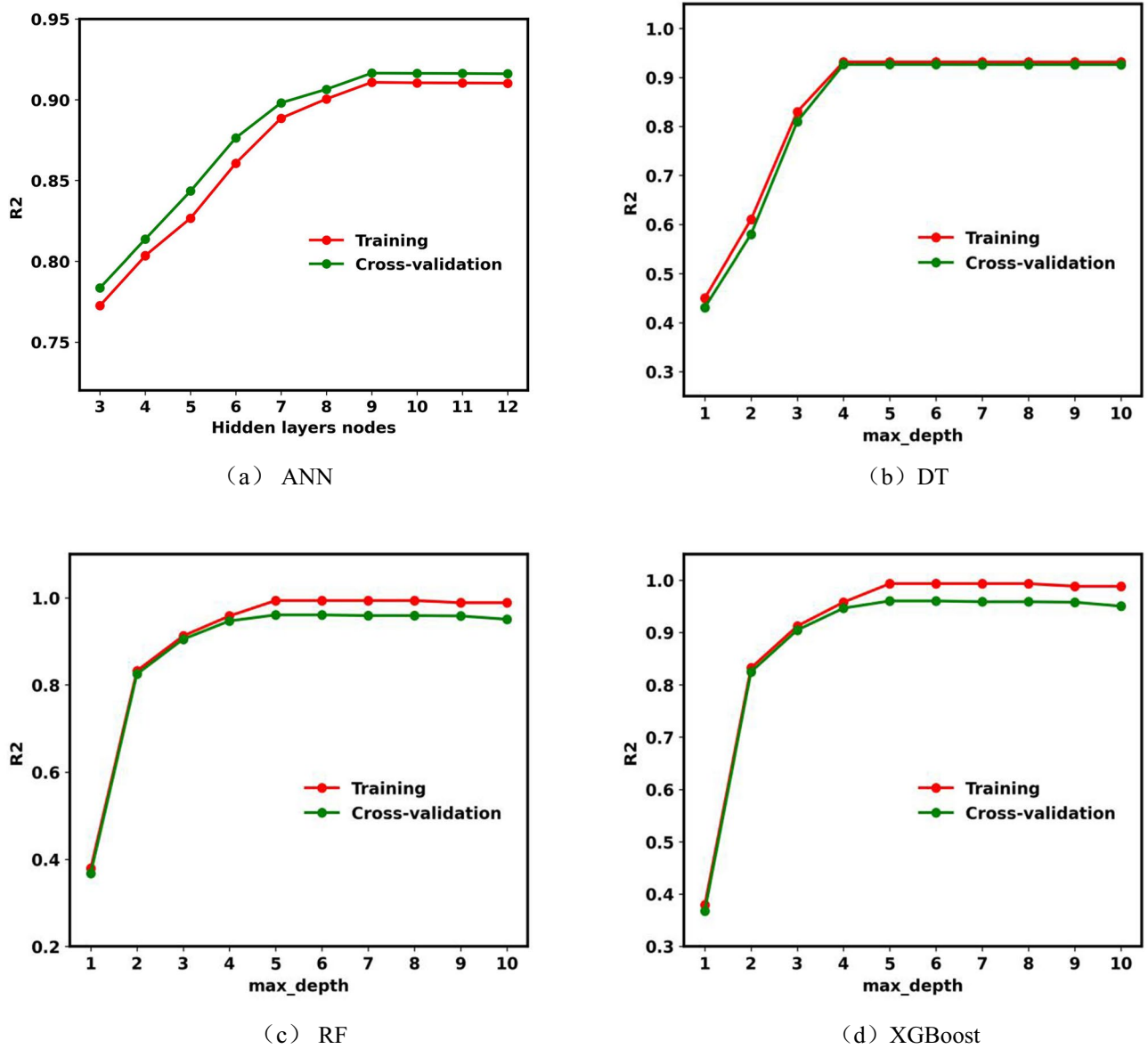**Table 3.** The search range of model hyperparameters and the final values of selected hyperparameters.

（a） ANN　　　　　　　　　　　　　　　　（b） DT

（c） RF　　　　　　　　　　　　　　　　（d） XGBoost

Figure 12. Validation curves using the critical hyperparameter.

Model	Training set			Test set		
	R^2	RMSE	MAE	R^2	RMSE	MAE
MLR	0.7489	0.0152	0.0113	0.7138	0.0134	0.0133
Decision tree	0.8965	0.0095	0.0068	0.8387	0.0117	0.0097
ANN	0.9108	0.0088	0.0055	0.9165	0.0086	0.0060
Random forest	0.9316	0.0078	0.0053	0.9265	0.0079	0.0065
XGBoost	0.9936	0.0024	0.0013	0.9605	0.0058	0.0041

Table 4. Comparison of predictive performance of machine-learning models.

Model generalization
The XGBoost model produced the best performances by far when compared to the other investigated models. In order to evaluate whether the XGBoost model demonstrated overfitting or underfitting, two evaluation metrics, namely RMSE and R^2, were used to evaluate the generalization of the proposed model in conjunction with cross-validation. The detailed results were shown in Fig. 14. Figure 14 illustrated the variations in RMSE and R^2 for XGBoost on both the training and validation sets. The RMSE reduced and R^2 increased in a fluctuating fashion

(a) MLR

(b) DT

(c) ANN

(d) RF

(e) XGBoost
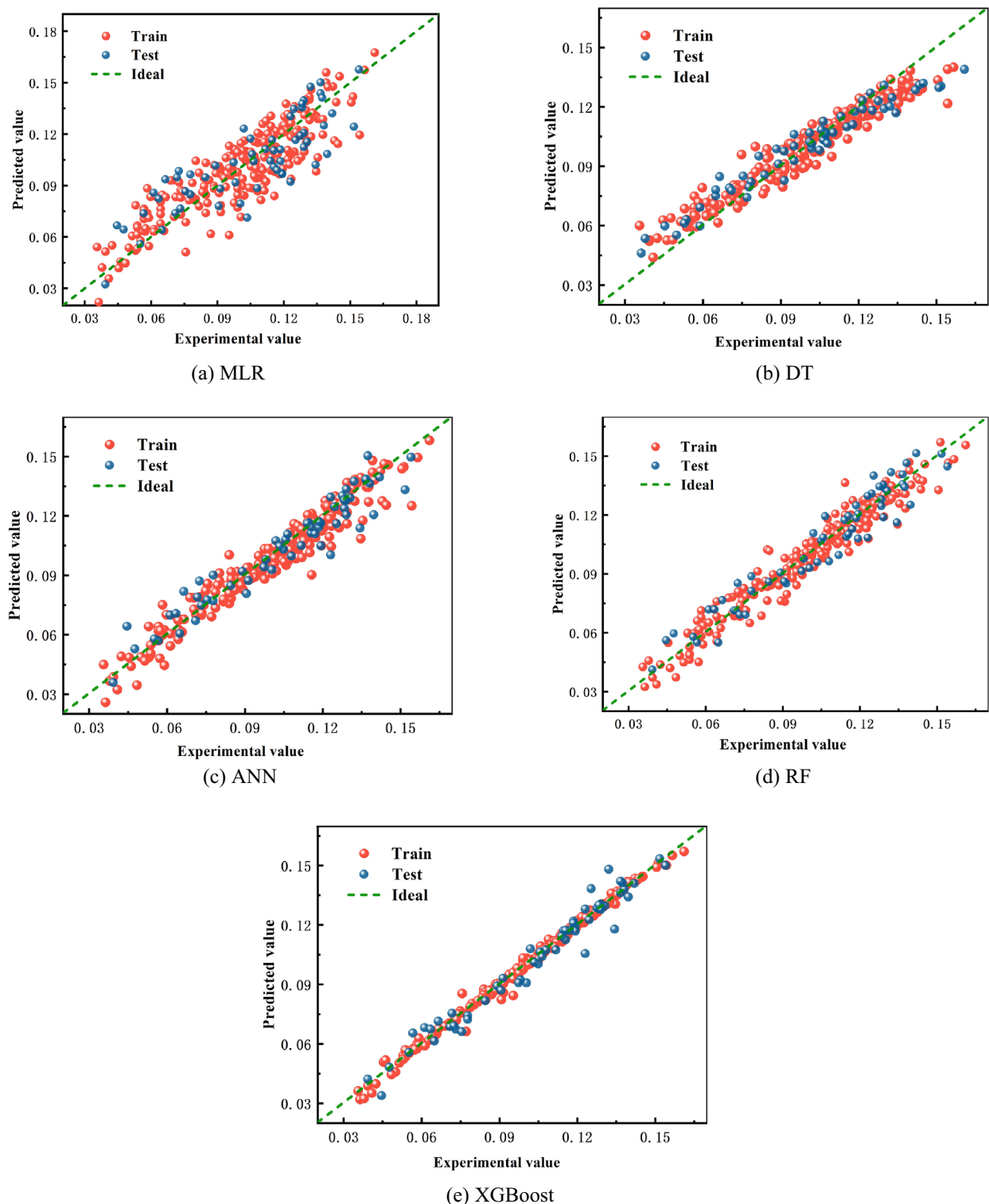
Figure 13. Fitting results of machine learning.

as the number of iterations increases. At the end of the experiment, both metrics tended to be constant, and the fitting performance on the validation set was close to that on the training set. This indicates that the XGBoost model did not suffer from overfitting or underfitting, demonstrating its good generalization.

To further demonstrate the robustness of the proposed model, predictive analyses were conducted on the 80 experimental datasets from reference[30] using Artificial Neural Networks (ANN), Random Forest, and XGBoost models. The prediction results were shown in Fig. 15 and Table 5. As indicated in Table 5, the XGBoost model proposed in this study demonstrated good performance in predicting the new dataset, achieving an R^2 of 0.9679,

(a) Loss function curve   (b) Accuracy curve
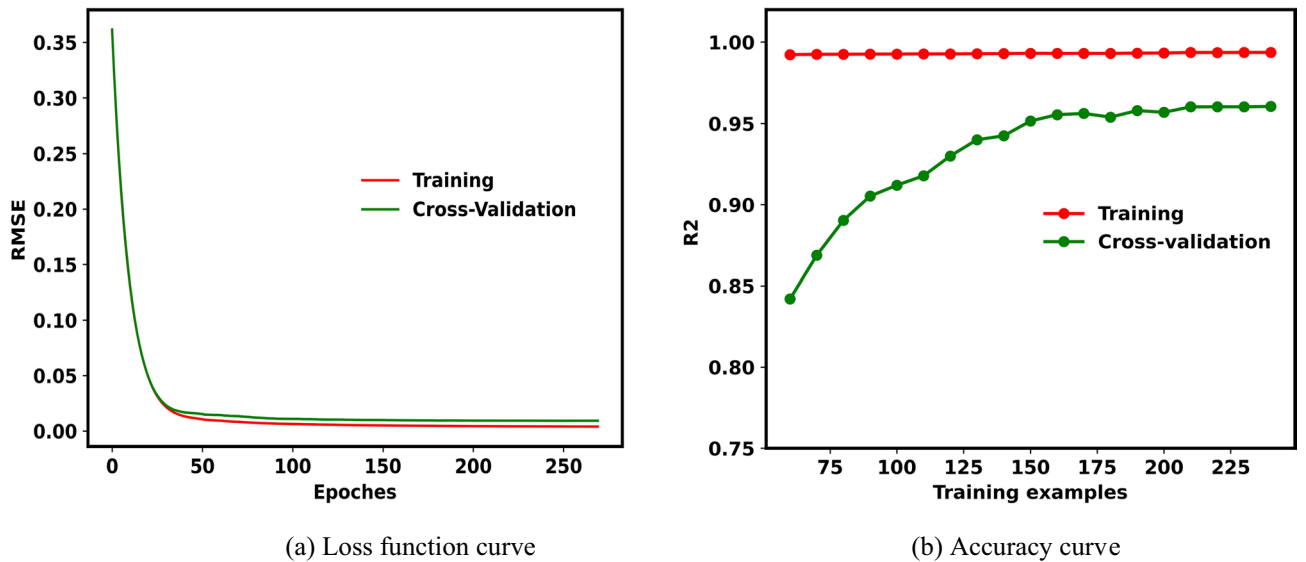
Figure 14. Overfitting analysis.
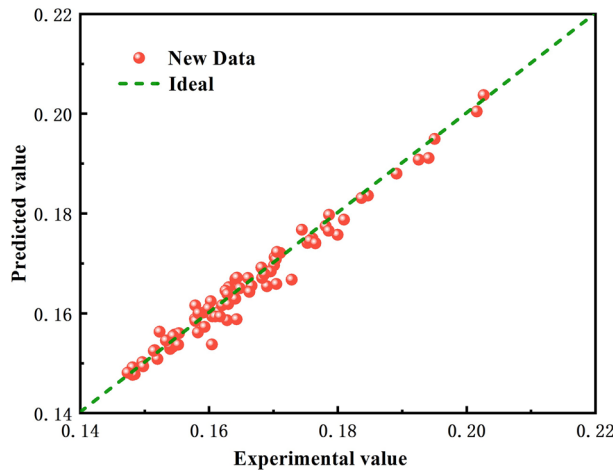
Figure 15. Predictions on the new dataset.

Model	Dateset	R^2	RMSE	MAE
ANN	Huttunen M[29]	0.9178	0.0035	0.0022
Random forest	Huttunen M[29]	0.9025	0.0038	0.0025
XGBoost	Huttunen M[29]	0.9936	0.0024	0.0013

Table 5. Comparison of machine learning model performance in predicting new data.

an RMSE of 0.0023, and an MAE of 0.0014. These favorable indicators demonstrated that the proposed XGBoost model exhibited high accuracy in predicting unknown data, and also indicated that the model possessed good generalization performance.

Dominant factor analysis of moisture content during vacuum screening
Correlation analysis of factors influencing moisture content
Before conducting the dominant factor analysis, it was necessary to analyze the correlation among six factors: the airflow rate, vacuum level, particle ratio, screen mesh, particle layer thickness, and residence time, as they all had varying degrees of influence on the moisture content of the filter cake. Correlation analysis helps with feature selection and managing multicollinearity, enhancing model performance and interpretability. This analytical step aids a deeper understanding of the data and provides a strong foundation for building more effective machine

learning models. The most widely used method of data correlation analysis is the Pearson correlation coefficient method. It quantifies the strength of the linear relationship between two continuous variables by calculating the ratio of their covariance to the product of their standard deviations, yielding a linear correlation measure that ranges from −1 to 1. The Pearson's correlation coefficient, when its absolute value is close to 1, suggests that there is a stronger linear relationship between the two variables. Conversely, when the absolute value is close to 0, it indicates a weak or negligible linear association. A heat map of Pearson's correlation coefficients of the input factors is shown in Fig. 16.

Values of $|\rho| > 0.4$ were observed for the airflow rate, particle ratio, and vacuum level (Fig. 16). This indicates that the airflow rate, vacuum level, and particle layer thickness exhibit a moderate level of linear correlation. Figure 17a is a scatter plot of the airflow rate and vacuum level; as the airflow rate increases, the vacuum level increases. Figure 17b is a scatter plot of the vacuum level and particle layer thickness; as the particle layer thickness increases, the vacuum level increases. The correlation coefficients ($|\rho|$) between the other three factors, that is, the particle ratio, screen mesh, and residence time, are lower than 0.2, indicating a weak linear correlation among these three factors. Overall, no strong linear correlation exists among the six factors. Therefore, when determining the dominant factors for moisture content during vacuum screening, all six factors can be simultaneously analyzed as input features.
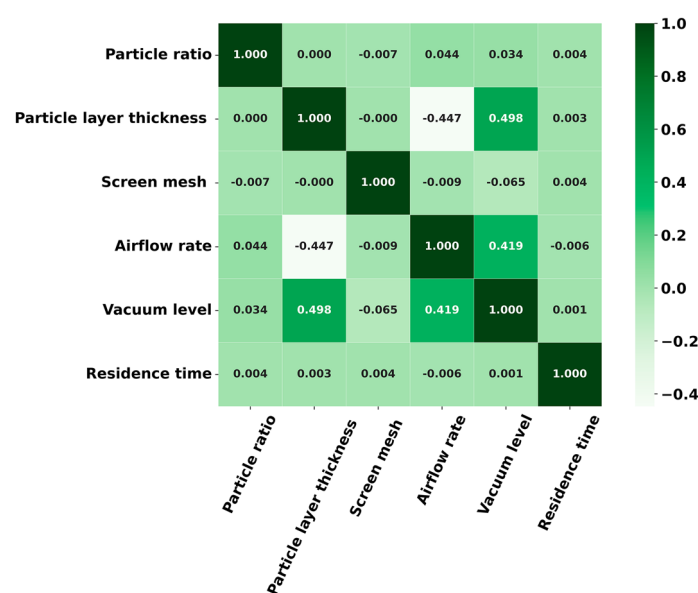

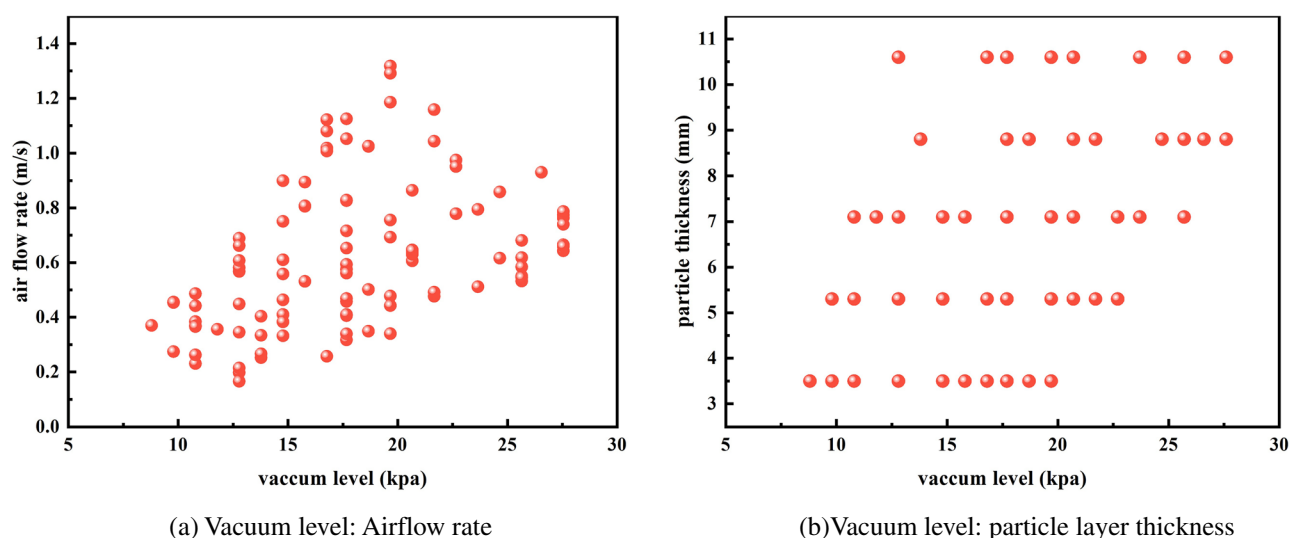**Figure 16.** Correlation coefficient heat map of input factors.

(a) Vacuum level: Airflow rate   (b) Vacuum level: particle layer thickness

Figure 17. Variable scatter plot.

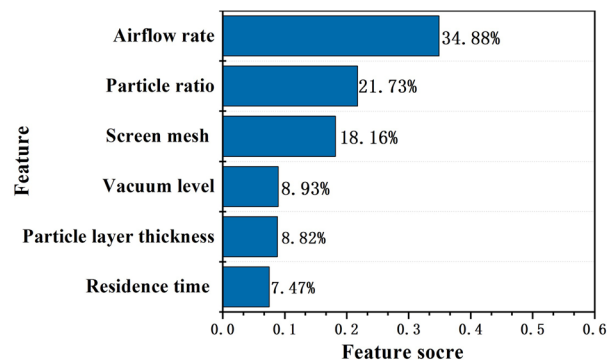## Identification of dominant factors for moisture content

The importance of machine-learning features can reflect the influence and contribution of the features to the target variable. Therefore, the dominant factors were evaluated using machine-learning feature importance. Three widely used methods for evaluating feature importance are tree-based feature importance, permutation feature importance, and SHAP[45]. Tree-based feature importance refers to the assessment of the significance of features within tree models, such as decision trees, random forests, and gradient-boosted trees, by quantifying the impact of each feature on the model's predictive outcomes. Permutation feature importance is a versatile approach that applies to any model, gauging the significance of features by inducing variability in their values and monitoring the subsequent effects on the model's predictive accuracy. SHAP is a game theory-based feature importance assessment method that decomposes the prediction results into the impact of each feature and assigns a SHAP value to each feature to represent its contribution to the prediction results. In this study, these methods were employed to analyze the dominant factors influencing the moisture content of the filter cake.

Figure 18 shows the feature importance rankings of the three prediction models. Figure 18a shows the ANN feature importance calculated using the permutation feature importance. The feature importance values of each factor ranged from 0.07 to 0.6. The top three features, in descending order of importance, were the airflow rate, particle ratio, and screen mesh. The contribution of these three features to the moisture content was higher than 20%, and the contribution of the remaining features was approximately 5%. Figure 18b and c depict the feature importance of RF and XGBoost, respectively, calculated using the tree-based feature importance method. The top three features had the same order as the ANN model (Fig. 18b), which are the airflow rate, particle ratio, and screen mesh. However, their contributions to moisture content differed from those of the ANN model, with contributions of 34.88%, 21.73%, and 18.16%, respectively. The last three features were the vacuum level, particle layer thickness, and residence time, with contributions of approximately 8%. The top three features were the particle ratio, airflow rate, and screen mesh (Fig. 18c). Their contributions to the moisture content ranged between 22 and 30%. The last three features were the vacuum level, particle layer thickness, and residence time, with contributions ranging between 5 and 13%.
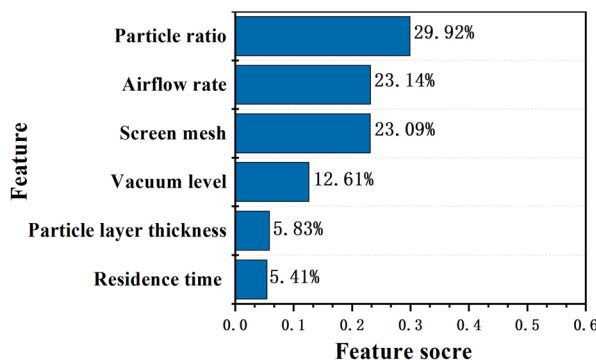
Overall, the feature importance rankings of the models differed. However, the three models were consistent in terms of the three most critical features, namely the airflow rate, particle ratio, and screen mesh, which had an average cumulative contribution of approximately 80% to the moisture content. The last three important features were the vacuum level, particle layer thickness, and residence time. These results indicate that the airflow rate, particle ratio, and screen mesh are the dominant factors influencing the moisture content during vacuum screening filtration, whereas the other three factors have less significant effects.
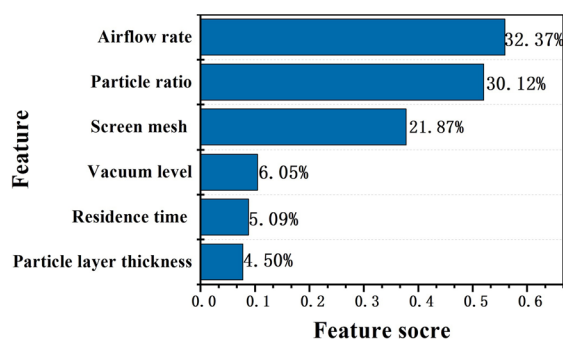
(a) Feature importance rankings by ANN

(b) Feature importance rankings by RF

(c) Feature importance rankings by XGBoost

Figure 18. Comparison of importance rankings of each model feature.

Interpretability analysis of dominant factors for moisture content

Based on the analysis presented in Section "Identification of dominant factors for moisture content", the dominant factors influencing the moisture content during vacuum screening were the airflow rate, particle ratio, and screen mesh. However, how these dominant factors influence the moisture content prediction results need to be clarified. In 2017, Lundberg and Lee developed a method for improving the interpretability of classification and regression models, namely, the SHAP model. The model combines the concept of the Shapley value in the cooperative game theory with the local interpretation method and constructs an interpretation framework for model prediction based on the feature contributions. The SHAP model not only reflects the degree of influence of each feature in each sample on the prediction results but also illustrates the positive and negative correlations of the influence.

A variant of SHAP (TreeSHAP) was used to interpret the model to determine the direction and magnitude of the relationship between the predictor and response variables. Figure 19 presents a summary plot of the feature importance rankings and the positive and negative impacts of the XGBoost model on the prediction of the filter cake moisture content and its outcomes in the SHAP framework. In this plot, the features are ordered from top to bottom based on their importance. Each point represents the SHAP value of a feature for prediction. All predictions were arranged along the horizontal axis based on the SHAP values and stacked along the vertical axis when the same value occurred to show the data distribution. The SHAP values reflect the positive and negative relationships between the features and prediction, and the color of the data points indicates the magnitude of the input feature values. As shown in Fig. 19, the features are ranked by their importance from high to low as follows: particle ratio, screen mesh, airflow rate, vacuum residence time, vacuum level, and particle layer thickness.

Interpretablility analysis of the effect of particle ratio
The most important feature was the particle ratio, which had a positive SHAP value and a positive correlation with the moisture content of the filter cake for samples with high values. A higher particle ratio indicates a higher proportion of fine particles in the mixture, which improves viscous and inertial resistance. This increased the resistance of the filtrate passing through the screen and filter cake, increasing the moisture content after vacuum-screening filtration.

Interpretablility analysis of the effect of screen mesh
The screen mesh was the second most important feature. The higher the feature value, the higher the moisture content of the filter cake, indicating a positive correlation between the moisture content and this feature. The particle ratio and screen mesh reflected the structure of the filter bed. Variations in the particle ratio and screen mesh indicated variations in the porosity of the filter bed. When the particle ratio increased, the proportion of fine particles in the mixed particles also increased, leading to an increase in the porosity of the filter bed. Similarly, increasing the screen mesh increased the porosity of the filter bed. As the porosity of the filter bed increased, the viscous and inertial resistances of the mixture also increased, increasing the resistance of the filtrate to passing through the particle layer during vacuum screening. This increased moisture content after filtration.

Interpretablility analysis of the effect of airflow rate
The airflow rate, which was the third most important feature, had higher SHAP values corresponding to samples with higher feature values, indicating a negative correlation with the moisture content after filtration. As the airflow rate increased, its effectiveness in transporting the pore water from the filter bed increased, decreasing the moisture content. This can be observed from the decreasing SHAP values with increasing feature values for the airflow rate.

These findings were consistent with the experimental results, further verifying that the particle ratio, screen mesh, and airflow rate were the dominant factors influencing the moisture content of the filter cake. Among the three dominant factors, the particle ratio and screen mesh reflected the filter bed structure during vacuum screening. Therefore, the optimization and application of vacuum filtration technology should pay more attention to the variation of airflow rate.
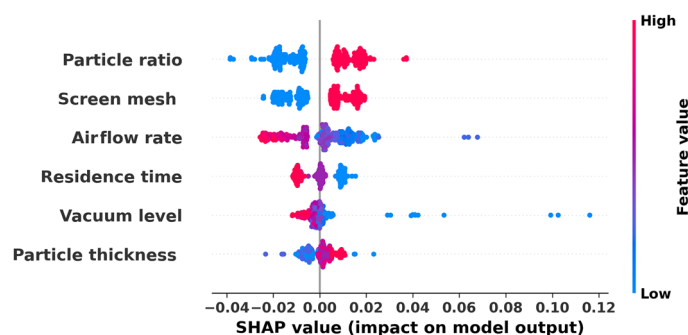

Figure 19. SHAP summary plot of input features.

Conclusions

In this study, a vacuum-screening apparatus was constructed, and experiments were conducted to predict the moisture content of the filter cake during vacuum screening. Three machine-learning methods were employed to establish the prediction model. The dominant factors influencing the moisture content were identified, and the following main conclusions were drawn.

(1) A vacuum screening experimental apparatus was constructed, and through the analysis of experimental data, it was found that the moisture content during the vacuum screening initially decreases rapidly and then more slowly, especially after the vacuum was sustained for 5 s, where the change tends to level off. This finding provided a direct basis for determining the optimal screening time, which was of significant importance for improving screening efficiency and reducing energy consumption.

(2) Five models, namely MLR, ANN, DT, RF, and XGBoost, were employed to establish moisture content prediction models for vacuum screening. The results demonstrated that the XGBoost model exhibited the highest prediction accuracy and stability, with a prediction accuracy of up to 96%,and the generalization of the model is verified by new unknown data. This provided a novel predictive tool for real-time monitoring and parameter tuning of the screening process, facilitating process optimization and efficiency. It was recommended that the XGBoost model and other machine learning techniques be applied to other types of screening processes to validate their generalisability and predictive capabilities.

(3) Through the SHAP value analysis and the comprehensive evaluation of five machine learning models, the particle ratio, screen mesh, and airflow rate were identified as the dominant factors influencing the moisture content during vacuum screening. This finding not only improves our understanding of the vacuum screening mechanism, but also provides a scientific basis for precise control of process parameters. It is suggested to focus on adjusting the particle ratio, screen mesh, and airflow rate to achieve better screening outcomes during the vacuum screening process.

Future research directions

(1) The effect of filtrate viscosity on moisture content during vacuum screening was overlooked in this study. The integration of filtrate viscosity as an important parameter[46] will be essential for a comprehensive evaluation of system performance in future.

(2) The sampling frequency of experimental moisture content should be enhanced to provide more robust data support in future work, which is essential for process optimization and control.

(3) Integrating machine learning and deep learning methods to predict the moisture content during vacuum screening may lead to more promising methods[47].

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Kroken, A., Vasshus, J. K., & Saasen, A., et al. A new fluid management system and methods for improving filtration and reducing waste volume, introducing a step change in health and safety in the mud processing area. In: *SPE/IADC Drilling Conference and Exhibition* (SPE-163522-MS, 2013).
2. Zhu, G. J. *et al.* Application of flocculant in vacuum filtration and dehydration of gold mine tailings. *J. Met. Mine* **4**, 45–52 (2021).
3. Wang, D. *et al.* Effects of mineral surface silanization and bitumen coating on its filtration from an aqueous slurry. *J. Fuel* **325**, 124921 (2022).
4. Rögener, F. Filtration technology for beer and beer yeast treatment. *IOP Conf. Ser. Earth Environ. Sci.* **941**(1), 012016 (2021).
5. Li, B. *et al.* Predicting the performance of pressure filtration processes by coupling computational fluid dynamics and discrete element methods. *J. Chem. Eng. Sci.* **208**, 115162 (2019).
6. Brownell, L. E. & Katz, D. L. Flow of fluids through porous media. *J. Chem. Eng. Progress.* **43**, 537–538 (1947).
7. Hoşten, Ç. & Sastry, K. V. S. Empirical correlations for the prediction of cake dewatering characteristics. *J. Miner. Eng.* **2**(1), 111–119 (1989).
8. Wakeman, R. J. Vacuum dewatering and residual saturation of incompressible filter cakes. *Int. J. Miner. Process.* **3**(3), 193–206 (1976).
9. Wakeman, R. J. The prediction and calculation of cake dewatering characteristics. *J. Filter. Sep.* **16**(6), 655–669 (1979).
10. Serajuddin, M., Anand Rao, K. & Sreenivas, T. Modelling and simulation of vacuum filtration of ore slurry: A case study on limestone-hosted Indian uranium ore. *J. Canad. Metall. Q.* **54**(4), 406–414 (2015).
11. Condie, D. J., Hinkel, M. & Veal, C. J. Modelling the vacuum filtration of fine coal. *J. Filtr. Sep.* **33**(9), 825–834 (1996).
12. Kerekes, R. J., McDonald, E. M. & McDonald, J. D. Decreasing permeability model of wet pressing: Extension to equilibrium conditions. *J-FOR* **3**(2), 46–51 (2013).
13. McDonald, J. D. & Kerekes, R. J. A decreasing permeability model of wet pressing. *Tappi J.* **74**(12), 142–149 (1991).
14. McDonald, J. D. & Kerekes, R. J. Pragmatic mathematical models of wet pressing in papermaking. *J. BioResources* **12**(4), 9520–9537 (2017).
15. McDonald, J. D. & Kerekes, R. J. Estimating limits of wet pressing on paper machines. *Tappi J.* **16**(2), 81–87 (2017).
16. Kerekes, R. J. & McDonald, J. D. Equilibrium moisture content in wet pressing of paper. *Tappi J.* **19**(7), 333–340 (2020).
17. Sjöstrand, B. *et al.* Numerical model of water removal and air penetration during vacuum dewatering. *J. Dry. Technol.* **39**(10), 1349–1358 (2021).

18. Rezk, K. *et al.* Modelling of water removal during a paper vacuum dewatering process using a Level-Set method. *J. Chem. Eng. Sci.* **101**, 543–553 (2013).
19. Li, W. *Research on Screening Mechanism of Negative Pressure Vibration Screen* (D. Southwest Petroleum University, Chengdu, 2018).
20. Lei, T. *Study on the Flow Law of Circulating Screen Mesh Negative Pressure Vibrating Screen Drilling Fluid* (D. Southwest Petroleum University, Chengdu, 2018).
21. Guo, F. *et al.* Coal gasification fine slag vacuum dewatering by ceramic membrane and numerical simulation. *J. Chem. Ind. Eng. Progress* **41**(8), 4047–4056 (2022).
22. Ma, W., Zeng, L., Zeng, Q., Zhang, S. & Wu, J. Numerical simulation and experimental verification of vacuum filtration. *J. Fluid Mach.* **50**(12), 49–55 (2022).
23. Liu, H. & You, K. Optimization of dewatering process of concentrate pressure filtering by support vector regression. *J. Sci. Rep.* **12**, 7135 (2022).
24. Gjelsvik, E. L., Fossen, M. & Tøndel, K. Current overview and way forward for the use of machine learning in the field of petroleum gas hydrates. *J. Fuel* **334**, 126696 (2023).
25. Ejerssa, W. W. *et al.* Loss of micropollutants on syringe filters during sample filtration: Machine learning approach for selecting appropriate filters. *Chemosphere* **359**, 142327 (2024).
26. Khan, M. A. *et al.* Application of random forest for modelling of surface water salinity. *J. Ain Shams Eng. J.* **13**(4), 101635 (2022).
27. Hakimi, M., Omar, M. B. & Ibrahim, R. Application of neural network in predicting H2S from an Acid Gas Removal Unit (AGRU) with different compositions of solvents. *J. Sens.* **23**, 1020 (2023).
28. Guerreiro, F. S., Gedraite, R. & Ataíde, C. H. Residual moisture content and separation efficiency optimization in pilot-scale vibrating screen. *J. Powder Technol.* **287**, 301–307 (2016).
29. Menezes, A. L. *et al.* Evaluation of the residual moisture content in pilot scale vibrating screening operating with pressure reduction in the screen drying region. *J. Powder Technol.* **369**, 17–24 (2020).
30. Huttunen, M. *et al.* Real-time monitoring of the moisture content of filter cakes in vacuum filters by a novel soft sensor. *J. Sep. Purif. Technol.* **223**, 282–291 (2019).
31. Jas, K. & Dodagoudar, G. R. Explainable machine learning model for liquefaction potential assessment of soils using XGBoost-SHAP. *J. Soil Dyn. Earthq. Eng.* **165**, 107662 (2023).
32. Homafar, A., Nasiri, H. & Chelgani, S. C. Modeling coking coal indexes by SHAP-XGBoost: explainable artificial intelligence method. *C. Fuel Commun.* **13**, 100078 (2022).
33. Alabdullah, A. A. *et al.* Prediction of rapid chloride penetration resistance of metakaolin based high strength concrete using light GBM and XGBoost models by incorporating SHAP analysis. *J. Constr. Build. Mater.* **345**, 128296 (2022).
34. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
35. Chunwen, Du. & Weibing, Z. Reasonable selection of basic parameters for drilling shale shakers. *J. Oil Field Equip.* **05**, 12–14 (2006).
36. Xianzhong, Yi. *et al.* Study on particle size distribution of drilling cuttings. *J. Pet. Mach.* **35**(12), 1–4 (2007).
37. Breiman, L. Bagging predictors. *J. Mach. Learn.* **24**, 123–140 (1996).
38. Ho, T. K. The random subspace method for constructing decision forests. *J. IEEE Trans. Pattern Anal. Mach. Intell.* **20**(8), 832–844 (1998).
39. Sagi, O. & Rokach, L. Explainable decision forest: Transforming a decision forest into an interpretable tree. *J. Inf. Fusion* **61**, 124–138 (2020).
40. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 785–794 (2016).
41. Abiodun, O. I. *et al.* State-of-the-art in artificial neural network applications: A survey. *J. Heliyon* **4**(11), e00938 (2018).
42. Hou, A. *et al.* Influence of variation/response space complexity and variable completeness on BP-ANN model establishment: Case study of steel ladle lining. *J. Appl. Sci.* **9**(14), 2835 (2019).
43. Pawlicki, M., Kozik, R. & Choraś, M. Artificial neural network hyperparameter optimisation for network intrusion detection. In *Intelligent Computing Theories and Application: 15th International Conference, ICIC* Nanchang, China, August 3–6, Proceedings, Part I 15, 749–760 (Springer International Publishing, 2019).
44. Tizakast, Y. *et al.* Machine learning based algorithms for modeling natural convection fluid flow and heat and mass transfer in rectangular cavities filled with non-Newtonian fluids. *J. Eng. Appl. Artif. Intell.* **119**, 105750 (2023).
45. Lundberg, S. M. & Lee S. I. A unified approach to interpreting model predictions. *J. Adv. Neural Inf. Process. Syst.* **30** (2017).
46. Agnihotri, J. *et al.* Higher frozen soil permeability represented in a hydrological model improves spring streamflow prediction from river basin to continental scales. *J. Water Resources Res* **59**(4), e2022WR033075 (2023).
47. Bahrami, B. & Arbabkhah, H. Enhanced flood detection through precise water segmentation using advanced deep learning models. *J. Civ. Eng. Res.* **6**(1), 1–8 (2024).

Author contributions

Funding

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to W.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.