FGHA-Net: Hybrid Attention Learning for Froth Flotation Recovery Prediction in Mineral Beneficiation

Mohammed I Thanoon[1], Fawaz Alanazi[2], Shaaban M. Shaaban[3]*, Asma A. Alhashmi[4],
Abed Saif Ahmed Alghawli[5], Ahlem Fatnassi[6], Majed Nawaz[7], Abdulbasit A. Darem[8]
Computing Department-College of Engineering and Computing at Al-Lith, Umm Al-Qura University, Saudi Arabia[1]
Department of Computer Science-College of Science, Northern Border University, Arar 73213, Saudi Arabia[2, 4, 6, 7]
Center for Scientific Research and Entrepreneurship, Northern Border University, Arar 73213, Saudi Arabia[3, 8]
Computer Science Department-College of Sciences and Humanities,
Prince Sattam Bin Abdulaziz University, Aflaj 11912, Kingdom of Saudi Arabia[5]

Abstract—**Froth flotation recovery prediction plays an important role in mineral processing because it can improve operational efficiency, reduce mineral loss, and support more effective decision-making during flotation operations. However, accurate recovery prediction remains challenging because flotation processes are highly dynamic and nonlinear, sensor data may contain noise, and process variables often exhibit complex temporal dependencies. Although several machine learning and deep learning approaches have been proposed for flotation prediction, many existing models remain limited in their capabilities of capturing short-term and long-range relationships in sequential flotation process data. To address these challenges, this study proposes a Fine-Grained Hybrid Attention Network (FGHA-Net) for froth flotation recovery prediction. The flotation process data are first preprocessed to improve data quality and suitability for model training. Feature engineering techniques, including process trend extraction, moving average features, and recovery-rate difference analysis, are then applied to enhance the representation of process dynamics. The proposed FGHA-Net consists of a bidirectional long short-term memory (BiLSTM) branch and a Transformer branch for hybrid temporal feature learning. The BiLSTM branch captures short-term sequential dependencies, whereas the Transformer branch models long-range temporal relationships using multi-head self-attention. A Fine-Grained Attention Fusion module is further introduced to support cross-branch feature fusion, temporal attention weighting, and refinement of important process variables. Finally, the fused features are passed to a regression head to generate the predicted recovery value. The proposed model is trained using the Adam optimizer and evaluated using standard regression metrics. Experimental results show that FGHA-Net achieved strong predictive performance on the testing set, with an RMSE of 0.011484 and an R^2 score of 0.844612. These results indicate that the proposed model can effectively learn complex temporal dependency patterns from froth flotation process data and has potential for intelligent mineral-processing monitoring and decision-support applications.**

Keywords*—*Froth flotation; mineral beneficiation; iron-ore processing; recovery prediction; soft sensor; BiLSTM; transformer; fine-grained attention; deep learning

I. Introduction

In mineral processing, froth flotation is one of the most widely used separation methods. Froth flotation is considered one of the most important and complex processes [1]. In froth flotation, concentrate grade plays a significant role in production performance. Online grade prediction is therefore essential for precise process control and effective optimization of industrial operations [2]. Because the direct measurement of key flotation variables often requires expensive instrumentation to acquire and maintain, soft-sensor modeling provides a fast and cost-effective approach for predicting concentrate grade based on variables measured online [3]. The predicted values can then be used as inputs to feedback control systems that adjust process variables to improve flotation performance. However, modeling froth flotation remains challenging because of its complex dynamics and strong interactions among process variables [4]. Existing models range from empirical approaches, such as artificial neural networks (ANNs), to phenomenological models, including kinetic, probabilistic, and population-balance models. Despite these approaches, no single model can fully capture the entire process dynamics, which often leads to considerable control challenges [5]. Data-driven approaches can identify hidden patterns and model nonlinear interactions; however, they still face several limitations. Machine learning (ML) involves the development of algorithms that enable systems to learn patterns from data and generate predictions without explicit programming [6, 7].

Effective preprocessing improves data quality and enables the model to capture relevant patterns and relationships [8]. Selecting an appropriate ML method is also essential because it strongly influences prediction accuracy, interpretability, and generalization ability. Different ML approaches have distinct strengths and limitations; therefore, choosing a suitable model allows the data to be represented more effectively and used for better decision-making [9]. Testing and validation are important for assessing model performance and ensuring that the model can generalize beyond the training samples [10]. The growing influence of ML and artificial intelligence (AI) is increasingly evident across various fields, from everyday applications to advanced scientific domains [11]. With rapid adoption in

*Corresponding author.

applications such as voice recognition, facial recognition, autonomous vehicles, social-media platforms, and spam filtering, these technologies are now widely used on a global scale [12]. In recent decades, researchers have identified several challenges in mineral beneficiation and have proposed potential solutions using ML- and AI-based models [13].

However, accurate froth flotation recovery prediction remains challenging because of process complexity, sensor noise, and strong dependencies among flotation variables. Existing ML and deep learning models also struggle to capture short-term and long-range temporal relationships effectively. Therefore, an efficient hybrid model is needed to improve flotation recovery prediction. Motivated by these challenges, this study proposes a Fine-Grained Hybrid Attention Network (FGHA-Net) for froth flotation recovery prediction. The proposed model combines BiLSTM and Transformer architectures to capture sequential temporal features and long-range dependencies from flotation process data. Additionally, a Fine-Grained Attention Fusion module is introduced to improve cross-branch feature fusion and important process-variable refinement. The contributing factors of this study are summarized as follows:

- This study integrates a bidirectional long short-term memory (BiLSTM) branch and a Transformer branch to capture short-term and long-range temporal dependencies from flotation process sequences.
- It introduces a Fine-Grained Attention Fusion (FGAF) module for effective temporal attention weighting and cross-branch feature fusion.
- It uses a regression head consisting of dense and linear output layers to generate the froth flotation recovery prediction output.
- It trains the proposed FGHA-Net model using the Adam optimizer to improve prediction performance and model convergence.

II. Related Work

This section presents a concise overview of existing studies on froth flotation recovery prediction and identifies the associated research gaps. Pandian et al. [14] proposed a deep probabilistic neural network technique to predict silica concentrate percentage and iron concentrate percentage in flotation processes. Their approach used a comprehensive flotation dataset that included reagent dosages, feed composition, column operating conditions, and pulp characteristics. Several machine learning models were evaluated, and a hybrid method combining a probabilistic pattern neural network with an Extra Trees regressor was proposed to improve prediction accuracy and effectiveness.

Zhang et al. [15] proposed an interpretable grade prediction model, which captures representative features from multi-source XRF sequential data. In addition, tensor reforming was applied to support efficient parallel evaluation. In [16], a convolutional long short-term memory neural network was proposed for real-time monitoring of chemical composition grades. The proposed implementation framework enables real-time monitoring and supports the adjustment of flotation parameters to improve process efficiency.

Yang et al. [17] investigated a coal flotation froth image segmentation method based on Mask R-CNN. Abouda et al. [18] proposed the AI-DeepFrothNet framework to address major limitations in froth image analysis. The framework used a Putrefaction Enrichment and Tuning Network (PETNet) to remove and reduce noise in froth images. It also applied multi-agent deep Q-learning, in which three agents were used to investigate different properties of the froth layer. In [19], a Gaussian Process-based Model Predictive Control algorithm was designed to incorporate uncertainty quantification directly into the control architecture. By exploiting the probabilistic nature of Gaussian Process modeling, this technique captures process variability and adapts dynamically to new information, enabling continuous refinement of the Gaussian Process model within the MPC framework. The Gaussian Process model was trained and optimized using JAX, evaluated with relevant performance metrics, and used as a surrogate model in the MPC approach.

Song et al. [20] introduced a prediction algorithm based on an improved XGBoost model using real copper concentrate flotation data. To address missing values and outliers in the collected dataset, the authors first proposed an outlier detection method. The integration of MissForest/IQR preprocessing with TPE-optimized XGBoost formed a sequential prediction pipeline for estimating copper concentrate grade in the flotation process. Bendaouia et al. [21] applied CNN-based feature extraction and image-processing techniques integrated with DL to predict the elemental composition of minerals. It included several machine learning methods, CNN models, and ANN models for predicting mineral concentrate grades. An overview of related studies on froth flotation prediction is presented in Table I.

Although several machine learning and deep learning methods have been introduced for flotation prediction and monitoring, several research gaps remain. Traditional regression and ML approaches have restricted ability in capturing complicated nonlinear and temporal dependencies, while existing DL models often inadequately model long-range relationships and complementary temporal representations. Moreover, limited attention has been given to adaptive feature interaction and fusion. To address these gaps, FGHA-Net integrates BiLSTM-based temporal feature learning, Transformer-based long-range dependency modeling, and fine-grained attention fusion to effectively combine complementary representations for improved flotation recovery prediction. Some image-based techniques mainly focus on froth image analysis; however, they do not fully exploit sequential process information for recovery prediction. Moreover, several recent approaches remain sensitive to noise, provide limited feature extraction capability, and are unable to capture both short-term and long-range process dependencies simultaneously. Therefore, an effective hybrid deep learning approach is still needed to learn temporal process patterns more efficiently and improve froth flotation recovery prediction performance.

TABLE I. SUMMARY OF RELATED WORKS IN FROTH FLOTATION PREDICTION

Reference	Aim	Methods Used	Input Data	Key Idea
Pandian et al. [14]	To predict silica and iron concentrate quality using probabilistic learning and ensemble-based regression methods	Deep Probabilistic NN and Extra Trees	Flotation process variables (reagents, pulp, operating conditions)	Hybrid probabilistic and ensemble learning
Zhang et al. [15]	To improve mineral grade prediction using interpretable RNN with shared parameters across grouped time series data	Grouped RNN with weight sharing	Multi-source XRF time series	Interpretable grouped sequential learning
Bendaouia et al. [16]	To monitor chemical composition grades by capturing both spatial and temporal patterns from flotation data	ConvLSTM	Video-based flotation data	Spatio-temporal feature learning
Yang et al. [17]	To extract froth bubble features and study their relationship with ash content in the flotation process	Mask R-CNN	Froth flotation images	Instance segmentation of froth bubbles
Abouda et al. [18]	To improve froth flotation monitoring by removing noise and analyzing froth conditions using deep reinforcement learning	DeepFrothNet + MA-DQL	Froth image data	Noise removal and reinforcement learning-based analysis
Dehon et al. [19]	To enhance flotation process control using uncertainty-aware Gaussian Process-based predictive control	Gaussian Process MPC	Process control data	Probabilistic model-based control
Song et al. [20]	To predict copper concentrate grade using an optimized XGBoost model with improved hyperparameter tuning	Improved XGBoost (TPE optimized)	Copper flotation dataset	Optimized tree-based prediction model
Bendaouia et al. [21]	To predict mineral concentrate grades using CNN-based feature extraction combined with machine learning models	CNN + ML hybrid	Mineral flotation dataset	Feature extraction from industrial flotation data

III. METHODOLOGY

The proposed model follows a systematic pipeline for froth flotation recovery prediction. The overall structure of the proposed framework is shown in Fig. 1. In the initial stage, data preprocessing is performed through missing-value handling, outlier removal using the interquartile range (IQR) method, Min–Max normalization, temporal sequence generation, and sliding-window formation. Next, feature engineering is conducted using process trend extraction, moving average features, and recovery-rate difference analysis. The proposed FGHA-Net then applies a dual-branch architecture consisting of a BiLSTM branch and a Transformer branch. A Fine-Grained Attention Fusion module integrates features from both branches by weighting and refining important process information. Finally, the fused features are passed to a regression head to predict flotation recovery, and the model is trained with the Adam optimizer.

A. Data Preprocessing

Data preprocessing is an important stage that improves data quality and prepares the dataset for model training. It includes missing-value handling, outlier removal, normalization, and temporal sequence generation [22].

Missing-value handling: A robust preprocessing pipeline is required to handle missing values in machine learning datasets. This step involves examining raw froth flotation process data to identify and correct missing values and inconsistencies. Removing low-quality values helps ensure that model performance is not affected by irrelevant, incomplete, or erroneous data.

Outlier removal using the IQR method: The interquartile range (IQR) method is used to identify and remove abnormal process values from the dataset. This method helps reduce the effect of extreme fluctuations and improves the consistency of the prediction model.

Data normalization: Normalization scales all attribute values into a specific range. Because feature values often vary across process variables, unscaled data may lead to unstable or inefficient training of deep learning models. Therefore, Min–Max normalization is applied to transform all values into the range between 0 and 1, which can accelerate model convergence and improve prediction performance, as shown in Eq. (1).

$$X' = \frac{X - \mathrm{Min}}{\mathrm{Max} - \mathrm{Min}} \quad (1)$$

In Eq. (1), X' represents the normalized feature value, while Min and Max denote the minimum and maximum values of the corresponding feature, respectively. Temporal sequence generation: The normalized flotation process variables are converted into temporal sequences to capture dependency relationships among process variables over time.

Sliding-window formation: A sliding-window method is used to divide the temporal sequences into fixed-length input windows for model training. This strategy enables the proposed FGHA-Net model to learn temporal dependency patterns from flotation process data.

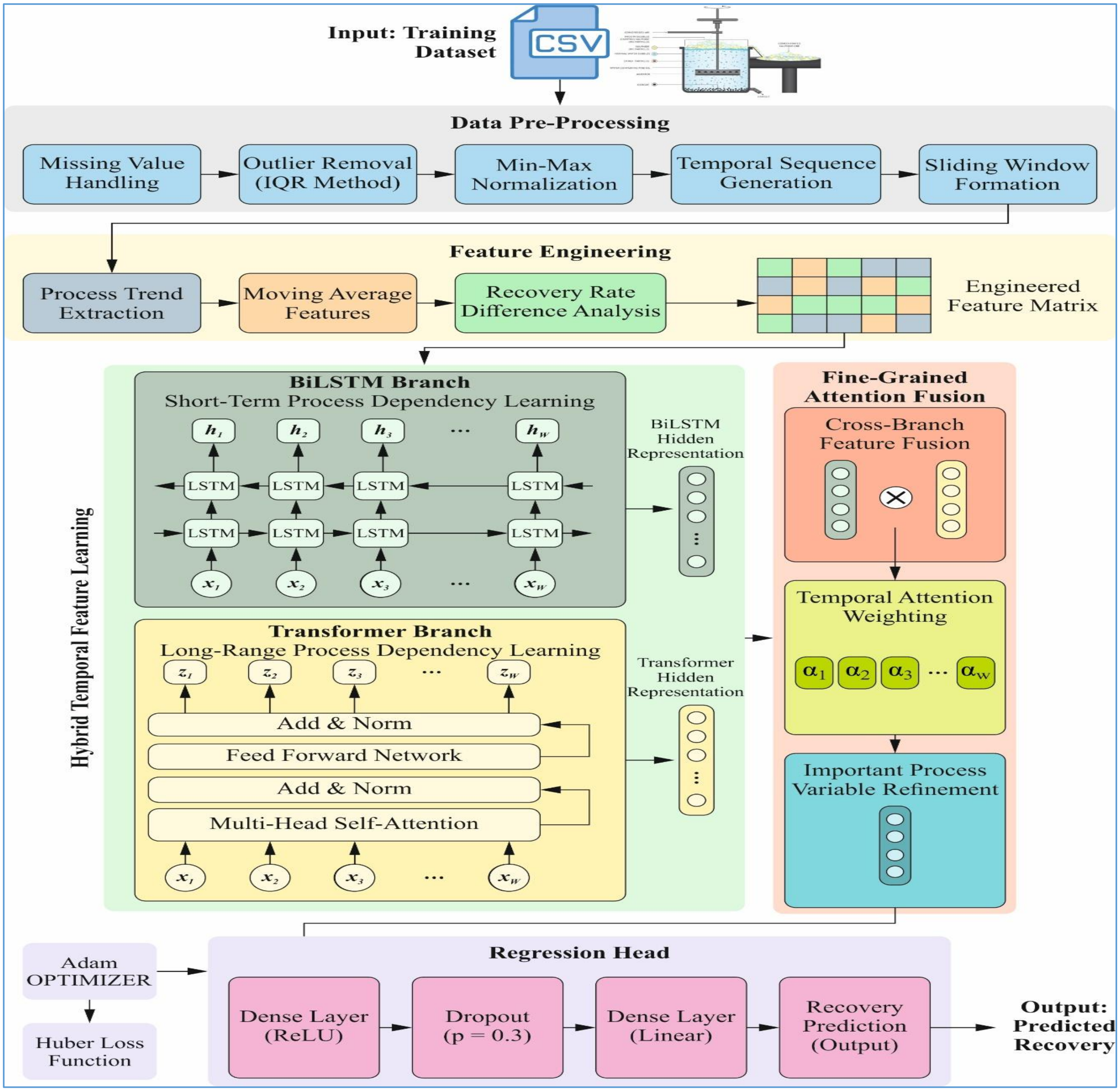

Fig. 1. Structural design of proposed framework.

B. Feature Engineering

Feature engineering is performed to improve the learning capability of the proposed FGHA-Net framework by extracting meaningful information from the froth flotation process data. Process trend extraction is used to capture the temporal variation of flotation process variables over time. These trend features help the model learn sequential dependency patterns and process changes. Moving average features are generated to reduce noise and smooth variations in flotation process sequences. These features help capture more stable temporal patterns from process variables. Recovery-rate difference analysis is performed to calculate the differences between consecutive recovery values and identify variability patterns in flotation recovery behavior.

C. Proposed FGHA-Net Architecture

The proposed hybrid temporal feature learning module is designed to capture both short-term and long-range dependencies in flotation process data. It integrates BiLSTM and Transformer branches to extract sequential and global temporal features effectively.

1) BiLSTM branch: The LSTM network consists of several key components, including the forget gate, input gate, output gate, and memory cell [23]. These components allow the network to retain relevant information and discard less important information over time. The LSTM network includes four structures: forgetting gate (f_i), memory unit, output gate (o_t), and input gate (i_i). The LSTM network holds additional cell states with triple gate elements in comparison with the RNN. In an LSTM structure, the forget gate (f_t) determines what percentage of data to retain in the network, and it is computed by:

$$f_t = \sigma\left(W_f \cdot [h_{i-1}, x_t] + b_f\right) \quad (2)$$

In this equation, h_{i-1} implies the preceding state memory, x_t represents the input sequence,$\cdot$ iV_f points out the weighted matrix of the forgetting gate, σ implies the sigmoid activation function, f_i signifies the state of the forget gate, and b_f denotes the bias of the forget gate. The input gate i_i is responsible for remembering the novel data in the cell state.

$$i_t = \sigma\left(W_j \cdot [h_{i-1}, x_t] + b_j\right) \quad (3)$$

$$\overline{C} = \tanh\left(W_c \cdot [h_{i-1}, x_t] + b_c\right) \quad (4)$$

$$C_i = f_i * C_i + i_i * \overline{C} \quad (5)$$

In these equations, C_i denotes the novel cell state, $\overline{C}$ refers to the cell state candidate, W_j denotes the weighted matrix of the input gate, tanh represents the hyperbolic tangent function, b_i signifies the bias of the input gate, iV_c implies the weighted matrix of the cell state, i_i specifies the input gate, and b_c shows the biased cell state, o_t denotes the output gate.

$$0_t = \sigma\left(|v_0 \cdot [h_{i-1}, x_t] + b_0\right) \quad (6)$$

$$h_t = 0_t * \tanh\left(C_t\right) \quad (7)$$

Here, (b_o) implies the bias of the output gate and (W_0)denotes the weighted matrix of the output gate.

This study uses a bidirectional LSTM (BiLSTM) branch, which processes sequence data in both forward and backward directions. The forward and backward hidden layers capture temporal dependencies from both past and future contexts. As a result, the BiLSTM branch extracts more informative sequential features and short-term process patterns, thereby improving the prediction accuracy of froth flotation recovery.

2) Transformer branch: The transformer branch is developed to obtain the global contextual data from sequences of the froth flotation method [24]. The transformer framework is constructed by the sequence embedding and multi-head self-attention mechanisms (MHSA) for encoding the temporal process data, and adaptably handling various measures of input signals, while modifying the computational complexity in the linear ranges. Thus, this branch is utilized as the encoder part for temporal feature extraction and long-term dependency learning. Likewise, the BiLSTM branch and the transformer module utilize the hierarchical feature extraction model to split the entire branches into numerous layers. At the 1st stage, the input processed sequences are primarily separated into temporal embeddings. Later, these embeddings can be entered into the transformer encoder model to obtain the 1st phase feature representations that are transferred to the following phase. The following feature representations are more effectively handled by employing the transformer encoder layers. In conclusion, the multi-scale temporal feature representations are gained and enable interrelationship amongst the layers of these two subdivisions.

To mitigate the higher computation complexity related to the self-attention mechanism (SAM) in transformer encoder parts, the transformer branch utilizes the MHSA to proficiently learn the dependency relations between the process variables. Same as the conventional attention mechanism (AM), the MHSA gains the value (V), query (Q), and key (K) as input, and generates improved temporal features.

$$MHSA\ (Q, K, V) = \text{Concat}\ (head_1, \ldots, head_{\text{n}})W^O \quad (8)$$

$$head_i = \text{Attention}\ (Q\text{W}_{\text{i}}^Q, KW_i^K, VW_i^V \quad (9)$$

Here the term ($head_i$) has the attention value $i - th$ attention head, the term Concat signifies the channel splicing function, and the parameters W_{i}^Q, W_i^V, and W_i^K indicate the linear projection parameters. The mathematical form of the attention operation can be measured by

$$\text{Attention}\ (Q, K, V) = Soft\ \max\left(\frac{QK^T}{\sqrt{d_k}}\right)\text{V} \quad (10)$$

Now, the factor (d_k) points to the key vector's dimensionality. By applying the MHSA module, the transformer branch successfully captures the long-term dependency correlation amongst flotation process variables, thus improving the ability of temporal representation and increasing the prediction efficiency of the proposed framework for froth flotation recovery prediction.

3) Fine-grained attention fusion: In the proposed framework, the cross-branch feature fusion includes convolution fusion, operator fusion, and attention fusion approaches [25]. The convolutional fusion approaches improve the cross-branch temporal features with respect to the semantic and multi-scale hierarchies by employing different convolution operations, decreasing the noise interference in the temporal features. Simple operator fusion techniques are directly added or concatenated to the features of the transformer and BiLSTM for fusion. The noise in the temporal features can be simply reduced by gaining dependable prediction outcomes by employing these techniques. Generally, the attention fusion methods utilize an AM to efficiently combine the cross-branch features in the temporal and channel dimensions. The convolutional techniques are focused on improving temporal features before fusion, whereas attention refinement techniques concentrate on boosting these features.

The FGAF model efficiently fuses features by employing the temporal attention data. The important part of the features is identified by employing temporal and channel attention, and the major parts amongst the BiLSTM and transformer branch features have been recognized by employing the temporal data. The channel branch accomplishes the combination of channel data through the transmission of input cross-branch features over the channels by employing global average pooling (GAP). The aggregation method is denoted by

$$fc = Concat(Avg(B_{\{input\}}, \text{Max}\left(B_{\{input\}}, Avg(T_{input}), \text{Max}\left(T_{input}\right)\right) \quad (11)$$

From this mathematical form, the parameter fc signifies the features from the fused channel. The Concat represents concatenation with the channel dimension. The GAP has been indicated by Avg Max ($B_{\{input\}}$ and T_{input}), which are the feature inputs of the BiLSTM and transformer branches, respectively.

The fused channel features can be entered into the 1D convolutions with effectively calculated kernel numbers (k) and efficiently obtain dependency among the channels. The formation of attention weights W_{c1} and W_{c2} in the channels employing the non-linear and softmax activation functions.

$$W_{c1}, W_{c2} = \frac{e^{Conv_1(f_c)}}{e^{Conv_1(f_c)} + e^{Conv_2(f_c)}} \frac{e^{Conv_2(f_c)}}{e^{Conv_1(f_c)} + e^{Conv_2(f_c)}} \quad (12)$$

Here, the terms $Conv_1$ and $Conv_2$ point to 1D convolution. This technique is employed for calculating the weights in the temporal and channel dimensions. Global pooling has been employed in the temporal dimension, and weights are acquired by the non-linear and convolution activation function. At the temporal branch, the significant method variable regions among the cross-branch features are computed by computing the temporal attention weights ($W_{\{T1\}}$) and ($W_{\{T2\}}$) parameters. Through the combination of temporal attention weights and cross-branch channel weights, the wide-ranging feature weights have been gained, thus emphasizing the vital parts amongst cross-branch features. Later, such feature weights can be multiplied by the cross-branch features for efficient fusion of these features. The final output must be handled by a 1×1 convolution and a convolution operation with the summation of the original inputs. The mathematical form of output is determined by:

$$f_{Boutput} = \text{f}^{conv}(f^{1\times1}(W_{C1} + W_{t1}) \otimes f_{Binput})) + f_{Binput} \quad (13)$$

$$f_{Boutput} = \text{f}^{conv}(f^{1\times1}(W_{C2} + W_{t2}) \otimes f_{Tinput})) + f_{Tinput} \quad (14)$$

From the above expression, the mathematical notation $\otimes$ denotes the multiplication operation, the parameter $f^{1\times1}$points to the set of ReLU activation functions, 1×1 convolutions, and batch normalization (BN), and the term f^{conv} refers to the convolution operation. By employing the proposed FGAF model, the model successfully performs temporal attention weighting, cross-branch feature fusion, and significant method variable improvement, thus increasing the prediction performance of the proposed FGHA-Net framework for froth flotation recovery prediction.

4) Regression head: A regression head is developed for froth flotation recovery prediction based on the FGHA-Net framework [26]. A dense layer with dropout regularization of 0.3 is fully connected using a regression head. This design choice stems from the necessity to effectively extract the refined temporal characteristics captured from the Transformer branch and BiLSTM branch.

This framework begins with a dense layer constructed to allow the fused temporal feature extraction. To reduce overfitting and improve the generalization capability, the extracted feature vector (h) is crossed over the dropout layer and dense layer. For the technique to operate a regression process for recovery prediction, a linear output layer can be applied to modify the last dense layer in Eq. (15). ($b_{\{out\}}$) and ($W_{\{out\}}$) indicates the bias and weight of the output layer.

$$y = W_{out}.h + b_{out} \quad (15)$$

Here (y) denotes the output of predicted froth flotation recovery. The regression head comprises trainable parameters related to the output and dense layers. This framework achieves the balance between computational efficiency and method intricacy. Within these integrated elements, the proposed FGHA-Net algorithm effectively functions to recover prediction outcomes for froth flotation process information. Fig. 2 represents the architecture of the hybrid temporal feature learning architecture combining BiLSTM and Transformer branches with fine-grained attention for prediction recovery.

D. Training Strategy

The proposed FGHA-Net model is trained using the Adam optimizer, which is an efficient stochastic optimization algorithm. Adam adaptively estimates learning rates for model parameters by combining momentum and adaptive gradient estimation [27]. This allows the optimizer to achieve stable and efficient convergence during deep learning model training.

In this study, the proposed FGHA-Net model is trained using Adam with a learning rate of 1e−4. Huber loss is used as the loss function because it is suitable for regression tasks and is less sensitive to outliers than mean squared error. The model is trained with a batch size of 32, and the dataset is divided into training and testing sets using a 70:30 split. The use of Adam optimization and Huber loss supports stable convergence and improves the robustness of the proposed recovery prediction model. This optimization technique is expressed as follows:

$$x_t = \delta_1 * x_t - 1 - (1 - \delta_1) * g_t \quad (16)$$

$$y_t = \delta_2 * y_t - 1 - (1 - \delta_2) * g_t^2 \quad (17)$$

$$\triangle\, \omega_t = -\eta \frac{x_t}{\sqrt{yt + \epsilon}} * gt \quad (18)$$

$$\omega_{t+1} = \omega_t + \triangle\, \omega_t \quad (19)$$

Here, the parameter g_t implies the gradient with respect to time (t), the symbol (θ) denotes the learning rate, the term x_t signifies the exponential averages of gradients, and the term y_t indicates the exponential averages of gradients squared, whereas δ_1 and δ_2 imply the factors.

To improve the strength of the predictive method, the proposed model was trained using the Adam optimizer with Huber loss as the loss function. The proposed system can be trained utilizing a batch size of 32 over 100 epochs. In addition, the database is separated using a 70:30 train-test split for performance evaluation and training. This method has revealed higher outcomes for regression-driven predictive tasks owing to the steady convergence action and adaptive parameter upgrading ability. These advantages make the Adam optimizer appropriate for the proposed FGHA-Net model for the froth flotation recovery predictive system.

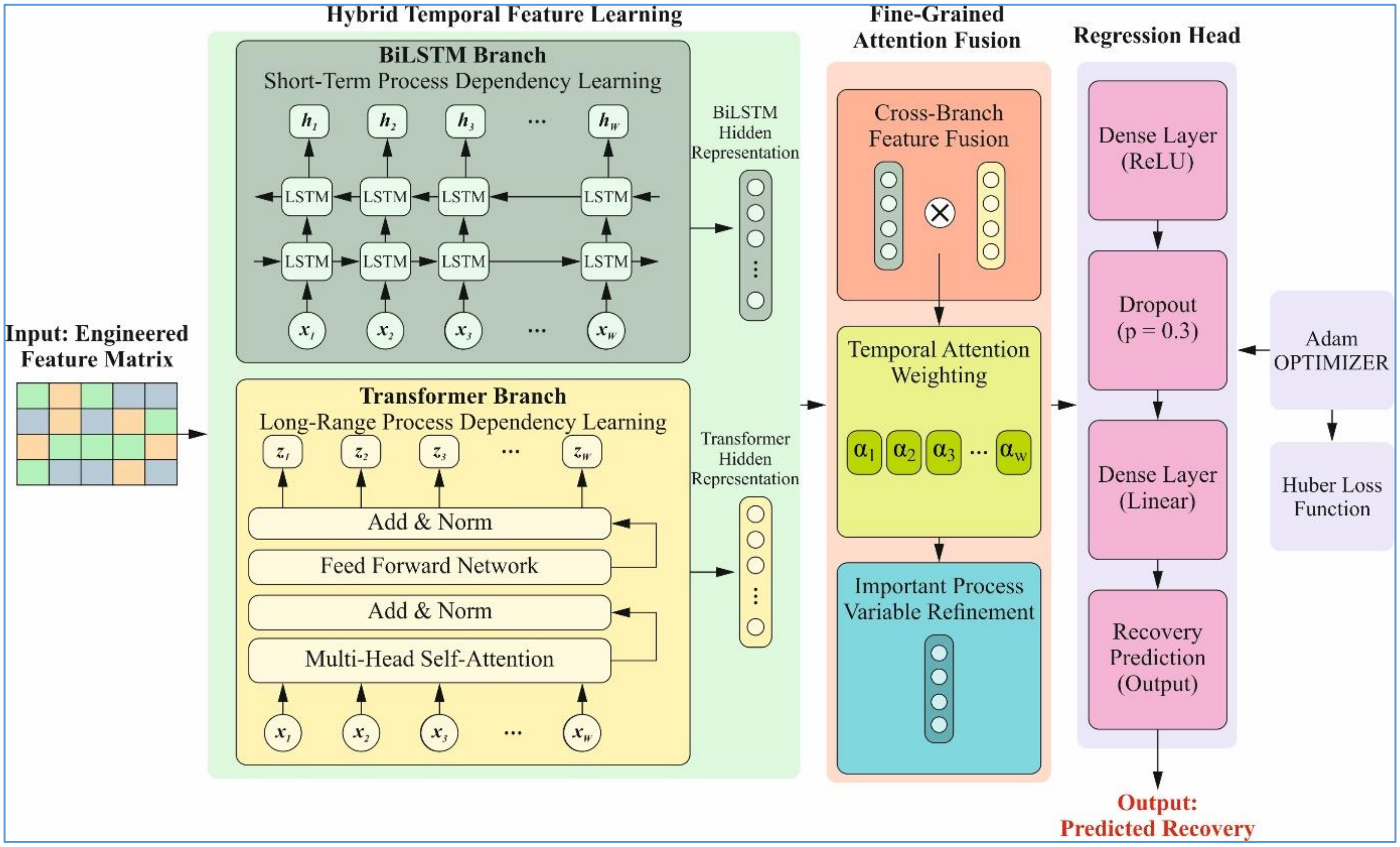

Fig. 2. Hybrid temporal feature learning architecture combining BiLSTM and Transformer branches with fine-grained attention for recovery prediction.

IV. Experimental Results and Discussion

The experimental evaluation of the proposed FGHA-Net model was conducted using a benchmark froth flotation dataset [28]. The dataset was collected from an industrial froth flotation plant. Column 1 contains date and time information, covering the period from March 2017 to September 2017. Columns 2 and 3 represent quality measurements of the iron ore pulp before flotation. Columns 4 to 8 include important variables that influence final ore quality. Columns 9 to 22 contain process variables, such as air flow and flotation-column level, which also affect ore quality. The last two columns represent laboratory-based measurements of the final iron ore pulp quality.

The dataset was divided into training and testing sets using a 70:30 ratio. Each input feature was normalized using Min–Max scaling. Temporal sequences were generated using a sliding-window technique, and the proposed FGHA-Net model was evaluated using standard regression metrics, including MAE, RMSE, MAPE, and R^2 score. The performance of the proposed model was also compared with several conventional machine learning and deep learning methods to evaluate its effectiveness.

Table II summarizes the parameter configuration of the FGHA-Net model. Feature engineering was performed using process trend analysis, a moving average window of 5, consecutive recovery-difference analysis, and 12 engineered features. The BiLSTM branch consisted of two layers, a dropout rate of 0.2, and 128 hidden units. The Transformer branch used four layers, eight attention heads, an embedding dimension of 128, and a feed-forward dimension of 256 with a multi-head self-attention mechanism. The Fine-Grained Attention Fusion module applied a cross-branch attention fusion strategy with 128 hidden units and an attention dropout rate of 0.2. The regression head consisted of dense layers with 128 and 64 neurons, followed by a linear output layer for recovery prediction. The model was trained using the Adam optimizer with a learning rate of 1e−4, Huber loss, a batch size of 32, and 25 epochs.

TABLE II. Parameter Settings

Model	Parameter	Values
Feature Engineering	Trend Extraction Method	Process Trend Analysis
	Moving Average Window	5
	Recovery Difference Metric	Consecutive Recovery Difference
	Engineered Feature Count	12
BiLSTM Branch	No. of BiLSTM Layers	2
	Hidden Units	128
	Dropout Rate	0.2

	Temporal Dependency Learning	Short-Term Sequential Learning
Transformer Branch	No. of Transformer Layers	4
	Attention Heads	8
	Embedding Dimension	128
	Feed Forward Dimension	256
	Self-Attention Type	Multi-Head Self-Attention
Fine-Grained Attention Fusion	Fusion Strategy	Cross-Branch Attention Fusion
	Attention Hidden Units	128
	Attention Dropout	0.2
	Feature Refinement	Temporal Attention Weighting
Regression Head	Dense Layer Structure	128 → 64
	Dropout Rate	0.3
	Output Layer	Linear
	Prediction Task	Recovery Rate Prediction
Training Configuration	Optimizer	Adam
	Learning Rate	1e-4
	Loss Function	Huber Loss
	Batch Size	32
	Epochs	25

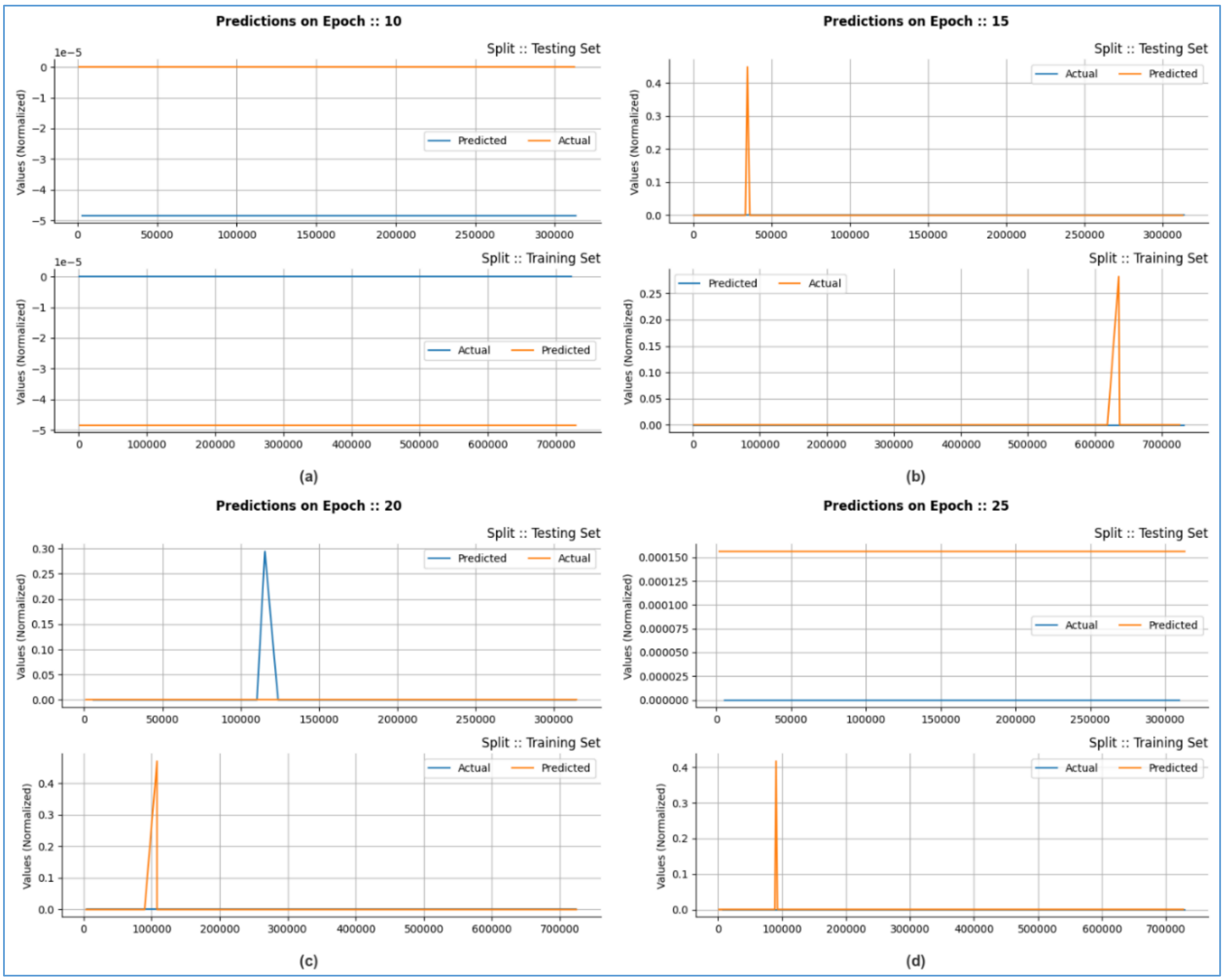

Fig. 3. Predictive analysis of the FGHA-Net method with different epochs.

Fig. 3 illustrates the prediction performance of the FGHA-Net model over different numbers of epochs, including 10, 15, 20, and 25 epochs, on both the training and testing sets. In the early epochs, the predicted and actual values show noticeable differences. However, as the number of epochs increases, the predicted values become more stable and show closer agreement with the actual values. These results indicate that the proposed model achieves stable convergence and improved prediction accuracy as training progresses.

Fig. 4 presents the loss-curve analysis of the FGHA-Net model on the training and testing sets over 25 epochs. The gradual reduction in MSE and RMSE indicates improved convergence and reduced prediction error. The MAE and MAPE results also demonstrate stable predictive performance. In addition, the increasing R^2 score indicates better model fitting and a stronger ability to learn flotation process behavior.

Table III and Fig. 5 present the performance evaluation of the FGHA-Net model using the training and testing sets. The model demonstrated consistent learning behavior, reduced prediction error, and stable convergence. On the testing set, the model achieved an MSE of 0.000132, RMSE of 0.011484, MAE of 0.000718, MAPE of 7.072856, and R^2 score of 0.844612. These results show strong agreement between the predicted and actual values and confirm the ability of the proposed framework to learn flotation process behavior effectively.

TABLE III. PERFORMANCE EVALUATION OF THE FGHA-NET MODEL WITH TRAIN-TEST SPLIT

Metrics	Training Set	Testing Set
MSE	0.000118	0.000132
RMSE	0.010878	0.011484
MAE	0.000683	0.000718
MAPE	7.006903	7.072856
R^2 Score	0.852408	0.844612

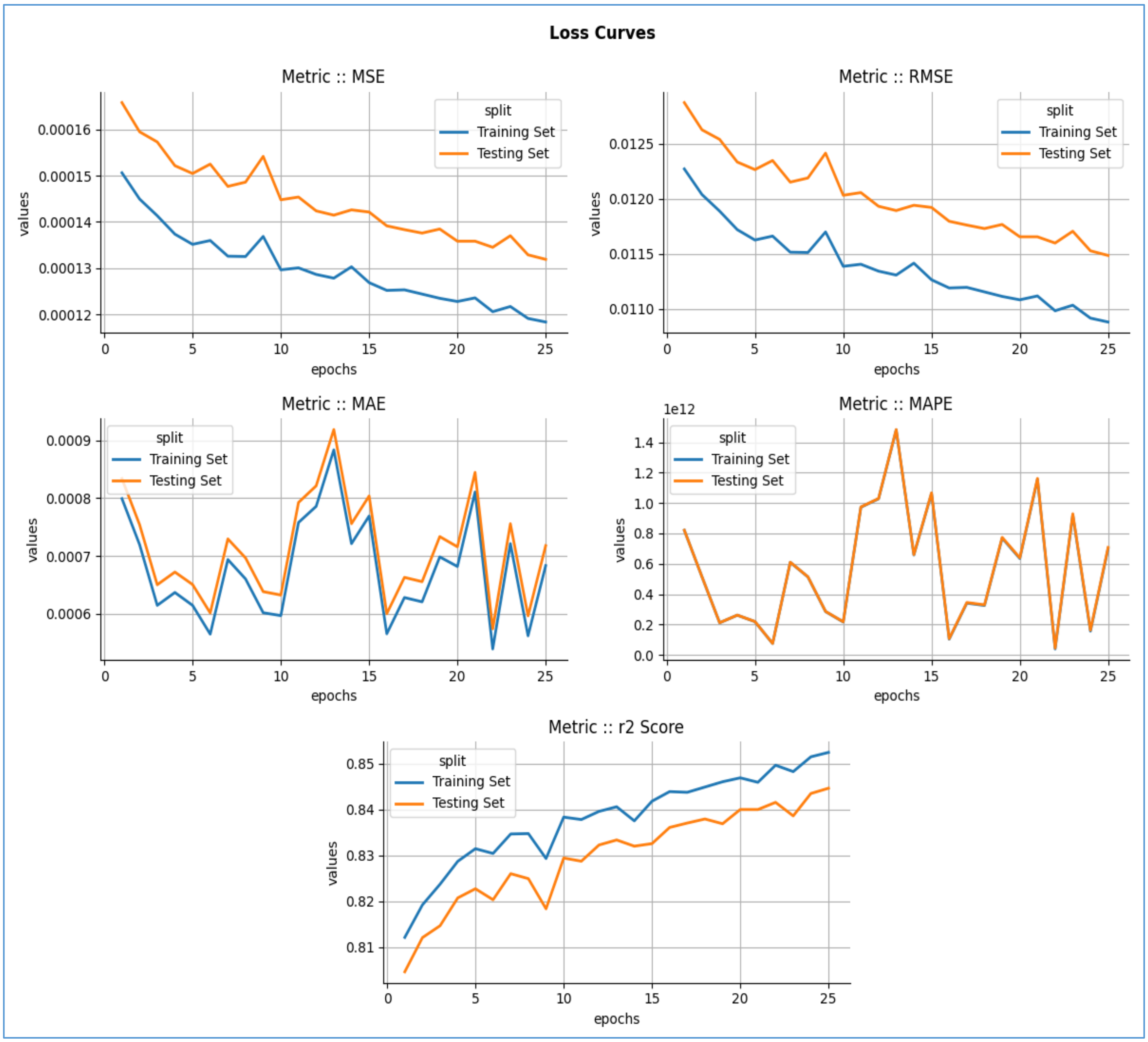

Fig. 4. Loss curve analysis of the FGHA-Net model.

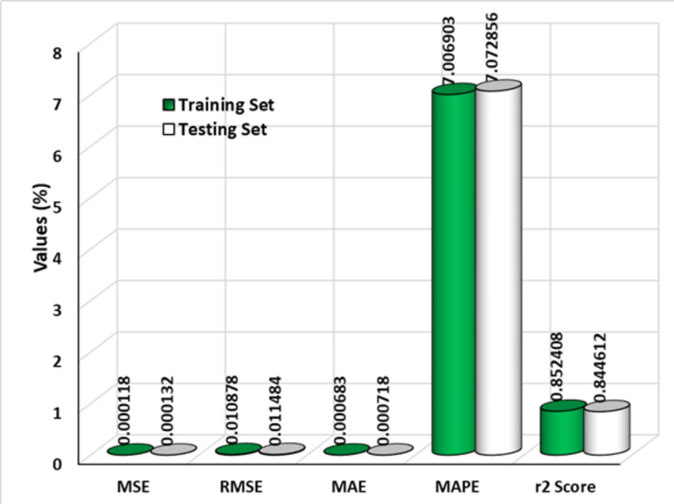

Fig. 5. Performance evaluation of the FGHA-Net model with training set and testing set.

Table IV presents the comparative assessment of the FGHA-Net model against existing methods for flotation recovery prediction [29-31]. Conventional models, such as Linear Regression, SVM, and M5P, achieved lower R^2 scores, indicating limited ability to capture nonlinear relationships. Other models, including FFNN, MLP-ANN, and Bi-PINN, showed moderate performance but still produced higher prediction errors than the proposed model. Concurrently, the hybrid models such as GA-NN, hybrid GA-NN, and BPN models have reached reasonable performance. But the proposed FGHA-Net outperforms existing models with minimal RMSE = 0.0109, MAE = 0.0007, MAPE = 7.0069, and R^2 = 0.8524, demonstrating better predictive performance than the compared models.

TABLE IV. COMPARISON ASSESSMENT OF THE FGHA-NET TECHNIQUE WITH EXISTING ALGORITHMS

Algorithms	RMSE	MAE	MAPE	R² Score
ANN	0.3120	0.0050	4.5900	0.4740
GA-NN	0.4000	0.0026	2.8800	0.7880
BPN	0.2590	0.0043	2.8700	0.2830
Hybrid-GA-NN	0.0568	0.0073	8.1200	0.5053
Linear Regression	0.0568	0.0073	8.1200	0.5053
FFNN	0.0552	0.0040	9.4380	0.6678
LSTM	0.0724	0.0025	13.7930	0.5662
Bi-PINN	0.0441	0.0069	11.8860	0.5024
MLP-ANN	0.0592	0.0043	8.8760	0.6143
SVM	0.0334	0.0058	12.1320	0.4621
RF-FFA	0.0825	0.0090	10.4920	0.5443
M5P	0.0229	0.0068	10.7300	0.6433
FGHA-Net	0.0109	0.0007	7.0069	0.8524

Table V provides the ablation study of the proposed model. The table values define the significance of each component involved in the proposed model. The experimental values imply that the proposed FGHA-Net (BiLSTM + Transformer + Fine-Grained Attention Fusion + Adam) gains enhanced performance over individual components.

Table VI portrays the comparative computational performance analysis of the proposed model in terms of FLOPs, GPU memory consumption, and inference time. The experimental outcomes demonstrate that the FGHA-Net accomplishes minimal computation requirements with 0.09 G FLOPs, 0.04 GB GPU memory, and an inference time of 0.20 ms, demonstrating its computational efficiency and suitability for real-time predictive applications compared with the baseline models.

TABLE V. ABLATION STUDY OF PROPOSED MODEL

Algorithms	RMSE	MAE	MAPE	R² Score
BiLSTM only	0.0141	0.0039	10.2269	0.5234
Transformer only	0.0135	0.0033	9.6869	0.5784
BiLSTM + Transformer	0.0128	0.0025	9.0169	0.6494
BiLSTM + Transformer (Without Fine-Grained Attention Fusion)	0.0120	0.0020	8.2769	0.7214
BiLSTM + Transformer + Fine-Grained Attention Fusion (Without Adam)	0.0115	0.0015	7.7569	0.7834
FGHA-Net (BiLSTM + Transformer + Fine-Grained Attention Fusion + Adam)	0.0109	0.0007	7.0069	0.8524

TABLE VI. COMPUTATIONAL EFFICIENCY ANALYSIS OF PROPOSED WITH EXISTING MODELS

Algorithm	FLOPs (G)	GPU Memory (GB)	Inference Time (ms/image)
ANN	0.24	0.35	6.8
GA-NN	0.41	0.48	8.4
BPN	0.32	0.41	7.5
Hybrid-GA-NN	0.68	0.72	11.3
Linear Regression	0.21	0.15	1.7
FFNN	0.32	0.43	7.8
LSTM	1.92	1.34	18.9
Bi-PINN	2.76	1.76	24.6
MLP-ANN	0.39	0.52	9.2
SVM	0.17	0.18	2.8
RF-FFA	0.15	0.29	4.3
M5P	0.13	0.08	1.2
FGHA-Net	0.09	0.04	0.20

The experimental results indicate that the proposed FGHA-Net model achieves lower error values and a higher R^2 score than the existing methods. This suggests that the predicted values are closer to the actual recovery values. The improvement can be attributed to the ability of the hybrid BiLSTM–Transformer architecture to capture both short-term and long-range temporal dependencies in flotation process data. The actual-versus-predicted plots further show that the proposed model captures variations in recovery more accurately than the compared methods. Overall, the results demonstrate that FGHA-Net is a promising approach for froth flotation recovery prediction.

V. Conclusion and Future Work

This study presented FGHA-Net, a Fine-Grained Hybrid Attention Network for froth flotation recovery prediction. The proposed approach first preprocesses the data to improve its quality and suitability for model training. Feature engineering techniques, including process trend extraction, moving average features, and recovery-rate difference analysis, are then applied to enhance the representation of flotation process dynamics. The FGHA-Net architecture consists of a BiLSTM branch and a Transformer branch for hybrid temporal feature learning. The BiLSTM branch captures short-term sequential dependencies, whereas the Transformer branch models long-range relationships using a multi-head self-attention mechanism. In addition, the Fine-Grained Attention Fusion module performs cross-branch feature fusion, temporal attention weighting, and important process-variable refinement. The fused features are then passed through a regression head to generate the recovery prediction output. The experimental results demonstrate that the proposed FGHA-Net model achieves promising prediction performance. On the testing set, the model obtained an RMSE of 0.011484 and an R^2 score of 0.844612, indicating low prediction error and strong agreement between predicted and actual recovery values. These findings suggest that the integration of BiLSTM, Transformer-based self-attention, and fine-grained attention fusion enables the model to capture both short-term and long-range dependencies in flotation process data. Despite these promising results, several limitations remain. Future work should evaluate the proposed model using larger real-time flotation datasets, additional process variables, and external industrial data collected under different operating conditions. Moreover, explainable AI techniques may be incorporated to identify the most influential process variables and improve the interpretability of the prediction results. The proposed model may also be integrated into real-time industrial monitoring systems to support practical mineral-processing applications. Future work can focus on robustness evaluation of the proposed model under sensor noise, missing measurements, equipment drift, and operational disturbances.

Acknowledgment

The authors extend their appreciation to Northern Border University, Saudi Arabia, for supporting this work through project number (NBU-CRP-2026- 289). This study is supported via funding from Prince Sattam bin Abdulaziz University, project number (PSAU/2026/R/1447).

Data Availability Statement

The data that support the findings of this study are openly available at https://data.mendeley.com/datasets/2vsrjbtfp5/1, reference number [28].

Conflicts of Interest

"The authors declare no conflicts of interest."

References

[1] Zhang, C., Tan, J., Peng, Y. and Xie, G., 2026. Froth image analysis for flotation process monitoring: a review. International Journal of Coal Preparation and Utilization, pp.1-41.

[2] Wang, H., Liu, Q., Xu, M., Xie, L., Xing, Y. and Gui, X., 2026. Semi-supervised clustering for operating condition recognition in sulfide ore flotation froth. Minerals Engineering, 237, p.110015.

[3] Tahooni Bonab, S., Abdollahi, H. and Abbaspour, A., 2026. A review of the mineral contact angle measurements and applications in the froth flotation process: Techniques and influential parameters, part II. Mineral Processing and Extractive Metallurgy Review, 47(1), pp.71-93.

[4] Leiva, C., Poblete, D., Humire, J. and Acuña, C., 2026. Kinetic analysis of froth flotation for remediation of hydrocarbon-contaminated coastal sands. Minerals Engineering, 244, p.110305.

[5] Lan, S., Dong, L., Shen, P., Mao, Y., Zhang, T., Zheng, Q., Liu, D. and Shen, Z., 2026. Green flotation depressats of typical salt and silicate gangue minerals in froth flotation–a comprehensive review. Green Chemistry.

[6] Kundu, T. and Parhi, P.K., 2026. Exploring the Flotation Performance of a Novel Low-Grade Indian Spodumene-Bearing Pegmatite Deposit for Sustainable Development. Journal of Sustainable Metallurgy, 12(1), pp.417-437.

[7] Liu, X., Feng, Y., Zhou, Z., Su, J., Wang, J., Zhang, Y., Wang, H., Jiang, H. and Wang, C., 2026. Nutrient-mediated surface hydrophobicity enhancing the removal of marine microplastics by froth flotation. Chemical Engineering Science, p.124080.

[8] Zhao, L., Zou, Z., Dong, B., Liu, L., Xie, Y. and Xie, S., 2026. Robust burst bubble recognition in flotation froths via RGB-D imaging and depth-guided feature extraction. Minerals Engineering, 242, p.110220.

[9] Barbouchi, A., Dachri, K., Labbilta, T., Naji, K., Faqir, H., Benzakour, I., Idouhli, R., Khadiri, M.E., Abouelfida, A. and Benzakour, J., 2026. Improving flotation recovery of gold in a refractory gold ore: An integrated approach between response surface methodology and automated quantitative mineralogy. Physicochemical Problems of Mineral Processing, 62(1).

[10] Benítez, J., Flores, V., Curilef, S., Martínez-Pelaez, R. and Leiva, C., 2025. Predicting copper recovery from flotation using machine learning and laboratory-generated data. Chaos: An Interdisciplinary Journal of Nonlinear Science, 35(9).

[11] Lu, M., He, X., Li, Y., He, L., Zou, Y., Yi, Z. and Li, P., 2025. Intelligent Optimization Setting of Froth Velocity by Integrating Multi-source Flotation Operation Information. IEEE Transactions on Automation Science and Engineering.

[12] Santos, E.G.D., Brum, I.A.S.D. and Ambrós, W.M., 2025. Techniques of Pre-Concentration by Sensor-Based Sorting and Froth Flotation Concentration Applied to Sulfide Ores—A Review. Minerals, 15(4), p.350.

[13] Yuan, M., Li, Q., Zhang, B., Pan, J., Gao, L., Jing, L. and Lu, L., 2026. A systematic review of deep learning methods for mineral exploration using multisource geoscience data (2018–2025). International Journal of Applied Earth Observation and Geoinformation, 149, p.105254.

[14] Pandian[1], P.K., Rohith, G., Kumar, R.L. and Harish, N., 2026. Deep Probabilistic Neural Network Model for Predicting Iron and Silica Concentrate Quality in Flotation Systems.

[15] Zhang, H., Yang, M., Tang, Z., Chen, J. and Xie, Y., 2026. Lightweight grade prediction in froth flotation using interpretable group Contribution-Enabled Weight-Sharing grouped RNN. Minerals Engineering, 237, p.110013.

[16] Bendaouia, A., Qassimi, S., Boussetta, A., Benzakour, I., Amar, O. and Hasidi, O., 2024. Artificial intelligence for enhanced flotation monitoring in the mining industry: A ConvLSTM-based approach. Computers & Chemical Engineering, 180, p.108476.

[17] Yang, S., Wen, Z., Zhou, M. and Zhou, C., 2026. An Innovative Extraction Method for the Morphological Features of Coal Flotation Froth Based on Instance Segmentation. Mining, Metallurgy & Exploration, pp.1-13.

[18] Abouda, K.A., Xu, D., Idress, W.M., Elmaki, H.M., Mazhar, T., Aoun, M., Ghadi, Y.Y., Shahzad, T. and Hamam, H., 2025. AI-DeepFrothNet: Continuous monitoring and tracking of froth flotation working condition by root cause analysis and optimized predictive control. Computers and Electrical Engineering, 123, p.110251.

[19] Dehon, V., Quintanilla, P., Chanona, A.D.R. and Van Impe, J.F.M., 2025. Probabilistic model predictive control for mineral flotation using Gaussian processes. Systems and Control Transactions, pp.1023-8.

[20] Song, Y., Yu, X. and Huang, M., 2025. A Copper Flotation Concentrate Grade Prediction Method Based on an Improved Extreme Gradient Boosting Algorithm. Applied Sciences, 15(20), p.11142.

[21] Bendaouia, A., Qassimi, S., Boussetta, A., Benzakour, I., Benhayoun, A., Amar, O., Bourzeix, F., Baïna, K., Cherkaoui, M. and Hasidi, O., 2024. Hybrid features extraction for the online mineral grades determination in the flotation froth using Deep Learning. Engineering Applications of Artificial Intelligence, 129, p.107680.

[22] Zhang, Y. and Wang, Z., 2023. Feature engineering and model optimization-based classification method for network intrusion detection. Applied Sciences, 13(16), p.9363.

[23] Hao, X., Liu, Y., Pei, L., Li, W. and Du, Y., 2022. Atmospheric temperature prediction based on a BiLSTM-Attention model. Symmetry, 14(11), p.2470.

[24] Xiong, X., Wang, X., Zhang, J., Huang, B. and Du, R., 2023. TCUNet: A lightweight dual-branch parallel network for sea–land segmentation in remote sensing images. Remote Sensing, 15(18), p.4413.

[25] Zhan, Z., Ren, H., Xia, M., Lin, H., Wang, X. and Li, X., 2024. Amfnet: Attention-guided multi-scale fusion network for bi-temporal change detection in remote sensing images. Remote Sensing, 16(10), p.1765.

[26] Gondaliya, K.M., Tsavdaridis, K.D., Raval, A., Amin, J.A. and Borisagar, K., 2025. Attention-Enhanced Progressive Transfer Learning for Scalable Seismic Vulnerability Assessment of RC Frame Buildings. Buildings, 15(23), p.4383.

[27] Mehmood, F., Ahmad, S. and Whangbo, T.K., 2023. An efficient optimization technique for training deep neural networks. Mathematics, 11(6), p.1360.

[28] Pu, Yuanyuan (2020), “Data for: FlotationNet: A hierarchical deep learning network for froth flotation recovery prediction”, Mendeley Data, V1, doi: 10.17632/2vsrjbtfp5.1, available at https://data.mendeley.com/datasets/2vsrjbtfp5/1

[29] Nasiri, M., Iqbal, S. and Särkkä, S., 2025. Physics-informed machine learning for grade prediction in froth flotation. Minerals Engineering, 227, p.109297.

[30] Cook, R., Monyake, K.C., Hayat, M.B., Kumar, A. and Alagha, L., 2020. Prediction of flotation efficiency of metal sulfides using an original hybrid machine learning model. Engineering Reports, 2(6), p.e12167.

[31] Allahkarami, E., Salmani Nuri, O., Abdollahzadeh, A., Rezai, B. and Maghsoudi, B., 2017. Improving estimation accuracy of metallurgical performance of industrial flotation process by using hybrid genetic algorithm–artificial neural network (GA-ANN). Physicochemical problems of mineral processing, 53.