Image-Based Iron Ore Feed Load Classification Using EfficientNet-B0 with Region-Density Attention and Hierarchical Feature Fusion

Sami F Karali[1], Fawaz Alanazi[2], Eman Ramadan Elsharkawy[3]*, Asma A. Alhashmi[4],
Abed Saif Ahmed Alghawli[5], Imen B. Mohamed[6], Chams Sallami[7], Abdulbasit A. Darem[8]
Computing Department-College of Engineering and Computing at Al-Lith 21691, Umm Al-Qura University[1]
Department of Computer Science-College of Science, Northern Border University, Arar 73213, Saudi Arabia[2, 4, 6, 7]
Center for Health Research, Northern Border University, Arar 73213, Saudi Arabia[3]
Computer Science Department-College of Sciences and Humanities, Prince Sattam Bin Abdulaziz University, Aflaj 11912, Kingdom of Saudi Arabia[5]
Center for Scientific Research and Entrepreneurship, Northern Border University, Arar 73213, Saudi Arabia[8]

***Abstract*—Iron ore estimation is a significant part of mineral exploration, as it supports determining both the quantity and quality of available mineral resources. Precise estimation facilitates effective mine planning, production management, and sustainable usage of natural resources. Conventional systems are mostly based on laboratory measurements and geological surveys that are frequently expensive, time-consuming, and restricted when handling larger-scale mineral datasets. In recent years, machine learning (ML) and deep learning (DL) systems have been progressively used to improve estimation by recognizing intricate patterns in mineralogical databases. This research introduces a Multi-modal Deep Learning Framework for Mineral Exploration and Iron Ore Load Prediction (MDL-MEOLP) model to enable precise mining decision-making. The proposed model performs hierarchical region generation utilizing a grid-driven patch extractor to capture multi-scale local regions. Subsequently, an EfficientNet-B0 backbone is employed for local feature extraction, while a region-density attention module emphasizes dense ore regions. Moreover, hierarchical feature fusion is applied to aggregate local-to-global representations and enhance global contextual understanding. Finally, a classification head with global average pooling, a dense layer, a dropout layer, and softmax activation is utilized to predict 17 feed load classes, and the model is trained utilizing Adam optimizer with cross-entropy loss over stratified data splits. To demonstrate the improved performance of the proposed MDL-MEOLP model, a comprehensive experimental analysis is carried out using Real-time Iron Ore Feed Load Estimation. The comparative results reported the effective performance of the MDL-MEOLP model with an accuracy of 97.58% over existing models.**

Keywords*—*Mineral exploration; Iron Ore Load Prediction; EfficientNet-B0; Adam optimizer; deep learning; resource utilization

I. Introduction

Over the past decade, the global need for mined resources has increased rapidly due to increasing population, advanced technologies, and economic growth [1]. The demand for metals will grow considerably by the year 2050, with a projected increase of 215% for aluminum, 140% for both copper and nickel, 86% for iron, 81% for zinc, and 46% for lead [2]. Metals like aluminum, copper, gold, and zinc play a vital role in the renewable energy transition and achieving net-zero targets because of their significant role in decarbonization technologies, including solar panels, wind turbines, and electric vehicles [3]. The growing demand for mineral resources has increased the pressure on the mining industry. Similarly, mining companies are facing numerous challenges like declining ore grades, deeper mineral deposits that are difficult to access, and rising energy and operational costs [4]. Meanwhile, society depends heavily on mineral resources for industrial growth and modern technologies, while mining itself is criticized for its environmental impacts [5]. Sustainable and transparent mining practices are vital to meet rising global demand while decreasing damage to the environment. Conventionally, the mining industry has relied on a linear value chain, including sequential stages from exploration to closure of the mine [6]. The procedure starts with exploration, in which all data is gathered to determine spatial extent, geometry, and geotechnical and hydrological settings of potential sites through various methods, including geological mapping, geophysics, remote sensing, and drilling [7].

In the mineral industry, the ball mill grinding circuit (GC) is a critical component for effective production, and low-grade iron ores must undergo a beneficial process for further concentration [8]. The product quality is mainly based on controlling several setpoints in the GC process, including feed load (ore load), water level, mill pressure, and pump pressure. The GC control practices are manually developed and depend on laboratory product analysis used as feedback for modifying those setpoints [9]. Among the various setpoints, the feed load setpoint is to be adjusted as the other adjustments need an extremely skilled and experienced human operator [10]. In simpler terms, the quality of the product is mostly influenced by the ore features and the feed load settings. For instance, when dealing with ore pellets that are difficult to grind, the feed load is kept low to maintain the quality of the product, while that is raised for easy-to-grind ore to improve product throughput [11]. With continuous grade control, precise ore characterization is possible and supports quick responses to variations in ore quality. In recent years, megatrends have emerged that could

*Corresponding author

significantly reshape the mining industry. Digital disruption, characterized by the incorporation of digital technologies into all operational functions, is transforming traditional practices [12]. The development of online sectors, Internet of Things (IoT) systems, and data analytics approaches, including artificial intelligence (AI) and geostatistical methods, enables efficient collection and evaluation of a large amount of real-time data. Such advancements support enhancing predictive maintenance, optimizing processes, improving efficiency, decision-making, and improving safety and sustainability [13]. AI-based technologies also support smart mining systems, which allow accurate ore deposit modeling, live monitoring, geo-metallurgical modeling, and continuous grade control. Together, these advanced technologies enhance resource utilization and minimize environmental impacts and also strengthen transparency in tackling the mineral resource dilemma [14].

A. Motivation and Research Contributions

In spite of developments in process control and mining operations, standard techniques are highly dependent on manual experience and the subsequent decision-making process, which constrains flexibility to change ore features. Current grinding circuit control methods often face difficulties in handling variations in mineral characteristics and feed load during real-time operations, which can reduce overall process efficiency. Although digital technologies and AI-based techniques have been introduced in recent years, many existing methods are still developed separately and are not fully integrated with practical operational control systems. Therefore, there is a strong need for an efficient data-driven framework that can improve ore classification and support accurate feed load prediction for better grinding circuit performance. To address these issues, the proposed MDL-MEOLP framework to design an efficient method for timely phases of mineral exploration employing iron ore images.

The key contributions of this study are outlined as follows:

- Construct a hierarchical region generation process employing grid-driven patch extraction to capture multi-scale local regions from iron images.
- Design an EfficientNet-B0-based feature extraction method for taking the fine-tuned texture patterns and ore density distributions from iron ore images.
- Introduce a region-density attention module to emphasize informative dense ore areas while suppressing irrelevant background data across adaptive feature refinement.
- Employ a hierarchical feature fusion approach to combine local and global representations, improving contextual understanding for feature representation.
- Apply a classification head with global average pooling (GAP), a dense layer, a dropout layer, and softmax activation to precisely predict 17 iron ore feed load categories.
- Integrate an Adam optimizer with cross-entropy loss to ensure stable training and enhance the performance of the method.
- Evaluate the performance of the proposed MDL-MEOLP utilizing the Real-time Iron Ore Feed Load Estimation database and exhibit the solution with several measures.

The proposed model's novelty lies in the domain-specific integration of its elements for the analysis of ore texture, rather than the elements themselves. It combines a grouped weight EfficientNet-B0 backbone, which explicitly divides local texture features from the ore density patterns, a Region Density Attention Module intended for the spatial sparsity of ore-bearing regions, which emphasizes dense regions and suppresses the sparse ones, and a hierarchical fusion pipeline which re-aggregates the multi-level regional characteristics in a closed loop with an attention refinement. This integrated model yields quantifiable gains over the generic backbone-plus-attention baselines utilized in prior ore classification systems.

The structure of this study is organized as follows: Section I explains the Introduction, and Section II reviews the related work on ore feed load estimation. Section III represents the proposed DL-MEOLP framework, Section IV describes the experimental validation and discussion, and Section V discusses the conclusion and future work of the study.

II. Literature Survey

This section discusses a brief overview of existing studies on iron ore feed estimation and identifies the key research gaps in the domain. Diniz et al. [15] developed a novel technique that applied a soft sensor algorithm to gather the energy efficiency of a wet-closed ball mill. Case research was shown through real information from one of the world's highest mining companies, VALE SA, using many models and validation methods. Matos et al. [16] designed a new data-driven soft sensor for evaluating the probabilistic distribution of multi-class lithology in the real world for crushing circuits. The methodology integrates measurements of crusher motor current and rotational speed with signal processing and ML models. Najafabadipour et al. [17] examined an innovative DL-based system to predict critical components in the Gohar Zamin iron ore mining region in southwest Kerman, Iran. One of the methods tested, the CNN, exhibited superior results in predicting the concentrations of the target component, attaining lower error rates and effectively extracting composite spatial patterns in the geochemical data. Firdaus et al. [18] investigated a strong generalized algorithm to automate the task of iron ore characterization of ores belonging to four diverse grades. Thus, a DL approach is designed, which employs a directed acyclic graph network framework through a hybrid inception topology.

Giannetti et al. [19] proposed an enhanced architecture for the prediction of silicon content by applying a new DL technique that depends on Phased LSTM. The system has been trained for three years of information and authorized under a one-year period, utilizing a strong walk-forward validation model, thus delivering confidence in the system results over time. Park et al. [20] discussed a new strategy to enhance the productivity of mining haulage mechanisms by deploying a hybrid methodology that integrates discrete event simulation (DES) and an ML system for predicting ore production. In [21], a hierarchical intelligent control model for mineral particle dimensions on the basis of ML is presented. During the ML layer, AI methodologies, namely CNN and LSTM are employed to resolve the multi-source ore blending prediction and intelligent categorization of rainy and dry season situations. Luo et al. [22] introduced a new Deep Multi-scale Spatial–Temporal Feature Decoupling Network (DMSTFD-Net), showing a clear attribute decoupling viewpoint to improve method interpretability.

A. Research Gaps

Numerous research gaps can be recognized on the basis of the relevant research. Diniz et al. [15] designed a soft sensor method for energy efficacy prediction in ball mills and emphasized that standard validation algorithms may cause misleading performance, demonstrating the requirement for appropriate time-series (chronological) validation. Matos et al. [16] developed a data-based soft sensor for lithology distribution in crushing circuits utilizing PLC-compatible signals; their strategy is still inadequate in generalization over multiple operational conditions and mining environments. Park et al. [20] examined ML with diverse event simulation for ore production and truck cycle time prediction, although real-world development in compound and extremely dynamic mining models remains difficult. This research shows that prediction accuracy has been enhanced; there is still a strong need for time-aware validation, the best method generalization over mines, and scalable real-world implementation in industrial mining classifications.

III. System Design

The manuscript follows a comprehensive pipeline for early-stage mineral exploration using iron ore images. To accomplish that, the proposed model initially preprocesses the input image through a median filter (MF), CLAHE-based enhancement, pixel normalization using min-max scaling, and resizing. In addition, the hierarchical region generation process is carried out by separating the iron ore image. Besides, the EfficientNet-B0 backbone is utilized to extract discriminative texture and ore density features from iron ore images. Moreover, the region-density attention module is applied to emphasize informative dense ore regions while suppressing less relevant background features. Hierarchical feature fusion is achieved to incorporate local-to-global representations for enhanced contextual understanding. Furthermore, the classification head with GAP, a dense layer, a dropout layer, and softmax activation is used to predict iron ore feed load classes. At last, the Adam optimizer is employed to ensure efficient model training and improved convergence. The complete workflow of the proposed model is illustrated in Fig. 1.

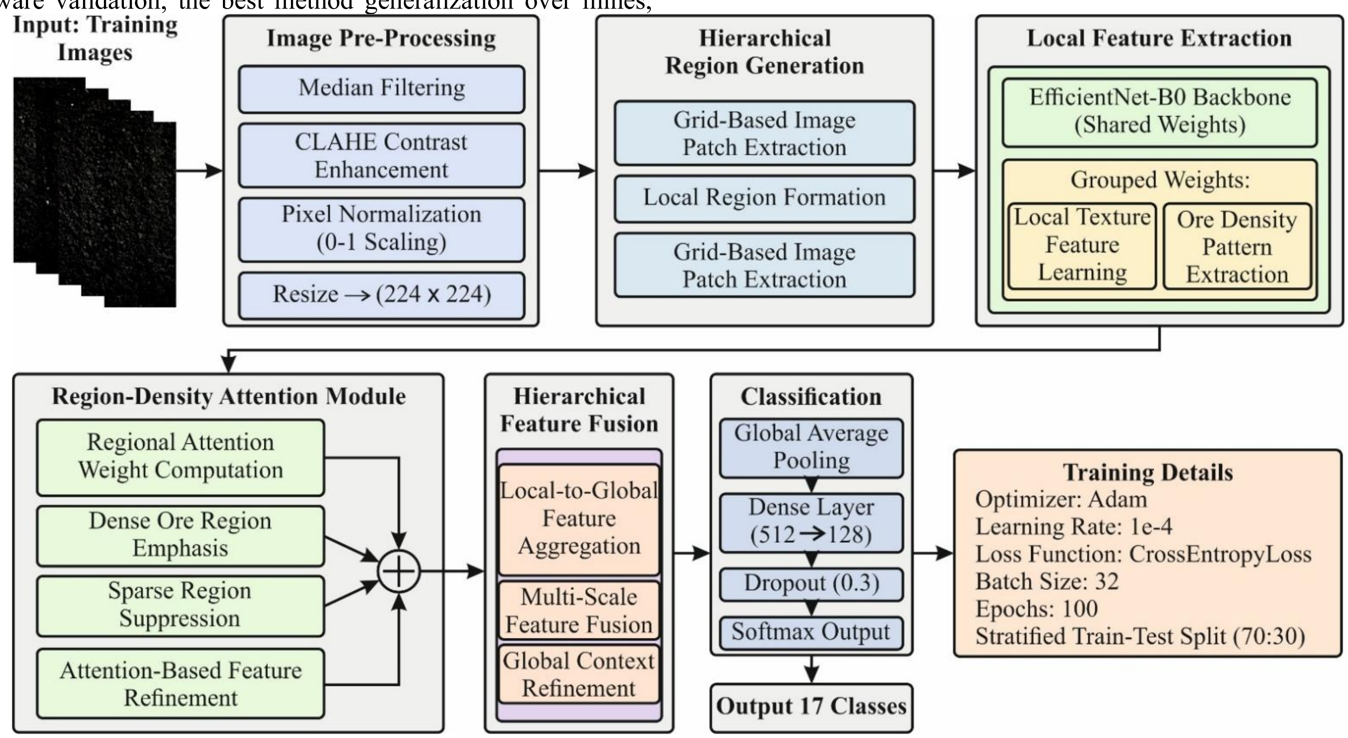

Fig. 1. Overall architecture of the MDL-MEOLP approach.

A. Image Preprocessing

Image preprocessing is an essential step as it improves image quality and ensures consistent input for better feature extraction. It initially preprocesses the input image through an MF, CLAHE-based enhancement, pixel normalization using min-max scaling, and resizing. The preprocessing pipeline consists of the following steps. Initially, the noise removal process is performed by median filtering for suppressing the sensor noise while preserving edge information, which is critical for texture analysis. Secondly, CLAHE is used for enhancing local contrast, to improve the visibility of fine-grained ore texture and variations in density. Third, all pixel intensities are rescaled by pixel normalization to the range of [0, 1] to stabilize and accelerate model training. Fourth, all the images are resized to a fixed resolution of 224 × 224 pixels to match the input needs of the EfficientNet-B0 backbone. Following the pre-processing, a hierarchical region generation step gets image patches by a grid-based scheme, forming local regions which are subsequently utilized for local to global feature learning.

1) Median filtering: To suppress salt-and-pepper noise while maintaining edges, MF is a non-linear image denoising method is utilized. It exchanges each pixel value with the median value of its nearby pixels in a definite window [23] [see Eq. (1)]:

$$I'(x,y) = median\{I(i,j) \big| (i,j) \in \Omega_{xy}\} \quad (1)$$

Here, Ω_{xy} implies the local neighborhood around pixel (x,y).

2) Contrast enhancement using CLAHE: It is an adaptive contrast enhancement model that applies histogram equalization over small image regions rather than the entire image, which improves local contrast in images. For preventing over-amplification of noise, it constrains the contrast enhancement to clip the histogram at a predetermined threshold and reorganizes the clipped pixels. To enhance fine structural details in low-contrast images while preventing noise amplification, this technique is specifically effective.

3) Min-max based pixel normalization: Min-max scaling transforms data values into a fixed range, typically between 0 and 1, by linearly rescaling the original values based on the minimal and maximal values of the data [see Eq. (2)]:

$$x' = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2)$$

x_{min} points out the minimal value of the attribute; x_{max} indicates the maximal value of the attribute, and x' represents the normalized attribute.

4) Resizing: To standardize input sizes before feeding data into DL algorithms, image resizing is applied. Each image is uniformly scaled or resampled to a fixed resolution of 224×224pixels during this process.

B. Hierarchical Region Generation

Hierarchical region generation is a structural process utilized for decomposing the mineral imaging into useful spatial units at several levels. The term Multi-modal refers to the various complementary feature modalities excavated from a single image input, instead of various external data sources. Specifically, it decomposes the image-derived representations into two distinct feature modalities, local texture and ore density pattern, through grouped-weight branches of the EfficientNet-B0 backbone. Furthermore, they are processed via the Region-Density Attention Module and integrated via hierarchical feature fusion over local, regional, and global levels. This use of multi-modal describes several learned feature representations from a single modality as multi-modal feature learning.

1) Grid-based image patch extraction: The input mineralogical imaging is separated into an even grid of smaller patches. All patches specify a fixed spatial zone of the imagery, assuring comprehensive and reliable coverage of the complete mineral surface.

2) Local region formation: Afterward, the patch extractor gathers neighborhood patches together to create local areas. These areas are useful geological micro-structures like mineral grains, texture transitions, and boundaries.

3) Multi-scale region representation: For capturing the differences in mineral structure at various levels of feature, local zones are more ordered into multi-scale representations.

C. Local Feature Extraction

Local feature extraction was conducted utilizing an EfficientNet-B0 backbone to train fine-grained texture patterns from the input images. This stage helps in capturing important ore density variations and local structural details for better representation of mineral characteristics.

1) EfficientNet-B0 Backbone: The EfficientNet framework has progressed to classify an efficient scaler approach for the CNN model and enhance prediction outcomes. The compound scaler approach constantly scales network depth, width, and resolution employing a certain set of scaling coefficients [24]. In this method, the EfficientNet-B0 structure is intended as the baseline method. The EfficientNet model comprises numerous versions from B0 to B7, while the superior versions involve improved parameters and enhanced learning ability. Growing network depth allows the method to train higher and more intricate patterns, whereas rising system width supports capturing fine-tuned features from geological and mineral imaging. This technique uses several convolutional layers with mobile inverted bottleneck convolution blocks for extracting effective features and representation learning in mineral deposit and iron ore load estimation tasks. The compound scaler approach is represented as in Eq. (3) to Eq. (7):

$$w = \beta^{\varphi} \quad (3)$$

$$d = \alpha^{\varphi} \quad (4)$$

$$r = \gamma^{\varphi} \quad (5)$$

$$s.t. \alpha. \beta^2. \gamma^2 \approx 2 \quad (6)$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \quad (7)$$

Here, w implies network width, d signifies the network depth, and r indicates network resolution, respectively, whereas α, β, and γ represent uniform scalar coefficients decided through the optimizer on the baseline technique. EfficientNet consistently scales each network size utilizing the compound coefficient ϕ, allowing effective learning of mineral distribution patterns and iron ore features from geological imaging, while preserving computational complexity.

2) Local texture feature learning: Local texture feature learning is used for capturing the fine-tuned mineral surface differences, grain limitations, and spatial irregularities from ore instances. The texture description is obtained by analyzing neighborhood pixel intensity interaction employing Local Binary Pattern (LBP) representation:

$$LBP(x_c, y_c) = \sum_{p=0}^{p-1} s(g_p - g_c) 2^p \quad (8)$$

In Eq. (8), g_c implies the pixel intensity of the center, g_p denotes the pixel intensity of the neighborhood, and $s(\cdot)$ indicates the thresholding operation.

3) Ore density pattern extraction: The ore density pattern extraction was performed using the mineral concentrate operation and the regional density difference in the ore structure [see Eq. (9)]:

$$\rho(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{9}$$

Let $\rho(x)$ imply the ore density estimation, K denotes the kernel operation, h represents the bandwidth, and x_i with respect to neighboring instance points.

D. Region-Density Attention Module

The region-density attention module is developed to enhance feature representation of minerals by adaptable attention on higher-density ore regions, while it controls the lower-information background areas [25]. This also increases spatial differentiation by assigning attention weights that are based on feature significance and local ore concentration.

1) Regional attention weight computation: Regional attention weights are calculated to measure the significance of all spatial regions, depending on the feature responses that could be extracted utilizing more images [see Eq. (10)]:

$$\alpha_i = \frac{\exp(W_a^T f_i)}{\sum_{j=1}^{N} \exp(W_a^T f_j)} \tag{10}$$

Here, the symbol α_i denotes the standardized attention weight for area i, f_i signifies the regional feature vectors, and W_a implies the trainable attention parameter.

2) Dense ore region emphasis: Dense ore regions have been underscored by improving the feature contributions proportionately to their attention scores, thereby confirming effective representation of mineral-rich regions [see Eq. (11):

$$F_i^{dense} = \alpha_i \cdot f_i \tag{11}$$

Here, the parameter F_i^{dense} signifies the improved feature representation of dense mineral zones.

3) Sparse region suppression: Sparse or non-mineral regions can be controlled by decreasing their feature influence by applying the attention weighting [see Eq. (12)]:

$$F_i^{sparse} = (1 - \alpha_i) \cdot f_i \tag{12}$$

It helps in eliminating the noise from background textures and distinct geological patterns. So, the model has been developed to be more robust to visual clutter and surface artifacts normally existing in real-time ore imaging.

E. Hierarchical Feature Fusion

Hierarchical feature fusion model combines the representations extracted with various abstraction phases to make the integrate feature space that improves mineral pattern and increases expectable robustness.

1) Local-to-global feature aggregation: The local features are captured from fine-grained mineral textures, while the global features are encoded in the total structural contexts.

$$F^{global} = \phi\left(\sum_{i=1}^{N} w_i F_i^{local}\right) \tag{13}$$

In Eq. (13), F_i^{local} represent the local feature vectors, w_i implies trained significance weights, and $\phi(\cdot)$ denotes the non-linear transformation.

2) Multi-scale feature fusion: The multi-scale fusion combines the feature extraction with various resolutions to obtain the macro-level geological patterns and micro-level mineral textures [see Eq. (14)]:

$$F^{MS} = \sum_{s=1}^{S} \alpha_s F_s, \text{where } \alpha_s = \frac{e^{z_s}}{\sum_{k=1}^{S} e^{z_k}} \tag{14}$$

where, the parameter F_s characterizes feature maps corresponding to scale s, and the mathematical notation α_s are adaptable attention weights calculated by employing the softmax function. It permits the model to dynamically highlight the more informative measures based on the intricacy of mineral textures. This fusion increases the strength against differences in imaging resolution, mineral density, and grain size.

3) Global context refinement: The global context refinement improves combined features through the integration of long-term dependencies and decreases the inappropriate background regions [see Eq. (15)]:

$$F^{refined} = F^{MS} \odot \sigma(W_g G) \tag{15}$$

In this mathematical equation, the symbol $\sigma(\cdot)$ points out the sigmoid gating function, the factor G indicates global context embedding, and the term W_g refers to the learnable projection matrix. The element-wise multiplication $\odot$ is performed as an attention gate that reinforces mineral-specific indications, whereas sparse background zone or removing noises. Hence, this enhancement phase majorly increases the feature variability in diverse geological contexts. Fig. 2 denotes the structural design of feature extraction and feature fusion model.

F. Classification Head

The classification head is used to convert extracted features into final predictions. It consists of GAP, a dense layer, a dropout layer, and a softmax activation to accurately predict 17 iron ore feed load classes [26].

1) Global average pooling: This pooling method is used after the convolutional feature extractor blocks to transform spatial feature mappings into a compressed feature representation. It minimizes every feature mapping into a distinct scalar performance by calculating the average of each spatial position, thus removing the essential for a flattened layer and considerably minimizing the count of trainable parameters.

$$GAP(x) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_{i,j} \quad (16)$$

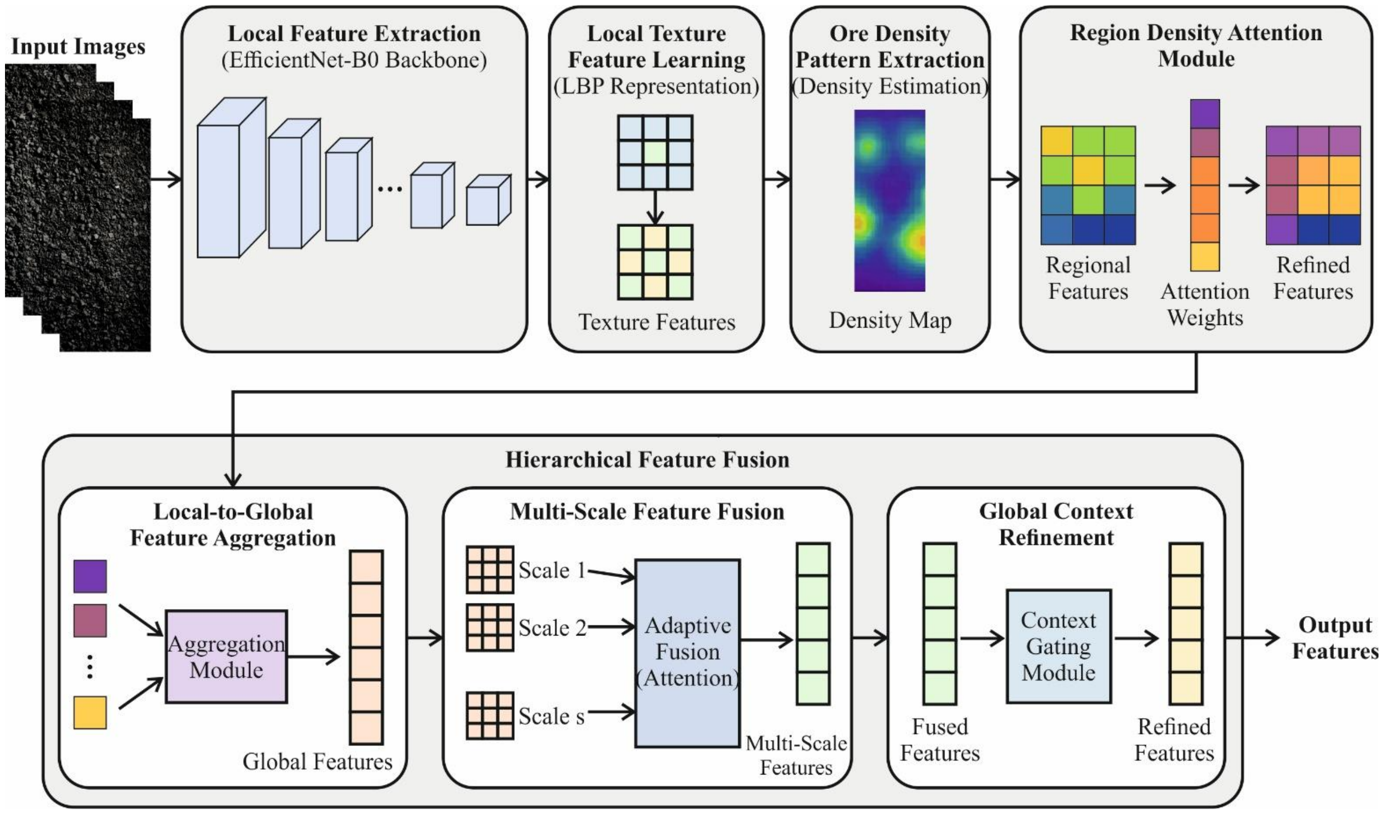

Fig. 2. Structural design of feature extraction and feature fusion model.

In Eq. (16), W and H represent the width and height of the feature mapping. This function considerably lessens the count of parameters compared with a flattened layer, leading to a more lightweight and computationally effective method while maintaining global spatial data.

2) Dense layer: This layer, also known as a fully connected layer, accomplishes non-linear feature alteration by mapping the dense feature vector from 512 to 128 neurons. It learns higher-level discriminative representations by using a weighted linear conversion followed by a non-linear activation function [see Eq. (17)]:

$$z = Wx + b \quad (17)$$

Here, x implies the feature vector of the input value, W implies the weighted matrix, and b denotes the biased vector.

3) Dropout: Dropout is a regularized method utilized to minimize overfitting for random deactivation neurons in training with probability p, forcing the system to learn stronger feature generalization [see Eq. (18)]:

$$y = Dropout(x, p) \quad (18)$$

4) Softmax output: The softmax function is used in the final layer to transform raw logits into a standardized probability function throughout numerous categories. It allocates high probabilities to further likely classes while assuring that the amount of each predicted probability equals to 1 [see Eq. (19)]:

$$softmax(x) = \frac{e^{z_i}}{\sum_j^4 e^{z_j}} \quad (19)$$

Now, z_i implies the logit of the i^{th} class, and K signifies the overall count of classes. In this study, $K = 17$ with respect to 17 feed load categorizations allows probabilistic multi-class identification for feed load estimation.

G. Network Configuration Details

Table I shows the parameter configuration of the MDL-MEOLP structure for feed load classification. Primarily, the input images are pre-processed employing Median Filtering, CLAHE-driven contrast enhancement, pixel standardization, and resizing images to 224×224 resolution to enhance the quality of imaging and consistency. Hierarchical area creation is performed using a grid-driven patch extractor with multi-scale local region representation, while EfficientNet-B0 is used as the backbone system to obtain density features and local texture. The structure integrates a Regional Density Attention module and hierarchical multi-scale feature fusion to increase significant global and regional contextual data. At last, the classifier procedure uses global average pooling and a Softmax layer for

17 feed load labels, and the method has been trained utilizing the Adam optimizer with a batch size of 32 over 100 epochs and a learning rate of 1e−4.

TABLE I. PARAMETER SETTING

Model	Parameter	Configured Value
Image Preprocessing	Filtering Method	Median Filtering
	Contrast Enhancement	CLAHE
	Normalization	Pixel Normalization (0–1 Scaling)
	Image Resize	224 × 224
Hierarchical Region Generation	Patch Extraction Method	Grid-Based Patch Extraction
	Patch Size	32 × 32
	Region Representation	Multi-Scale
	Number of Local Regions	16
Local Feature Extraction	Backbone Network	EfficientNet-B0
	Feature Extraction Type	Local Texture and Density Features
	Feature Embedding Dimension	1280
	Activation Function	Swish
Region-Density Attention Module	Attention Mechanism	Regional Density Attention
	Attention Hidden Units	256
	Sparse Region Threshold	0.2
	Attention Dropout	0.2
Hierarchical Feature Fusion	Fusion Strategy	Local-to-Global Aggregation
	Fusion Type	Multi-Scale Feature Fusion
	Global Context Refinement	Enabled
	Fusion Output Dimension	512
Classification Head	Pooling Layer	Global Average Pooling
	Dropout Rate	0.3
	Output Layer	Softmax
	Number of Classes	17 Feed Load Classes
Training Configuration	Optimizer	Adam
	Learning Rate	1e-4
	Loss Function	CrossEntropyLoss
	Batch Size	32
	Epochs	100
	Train-Test Split	80;20 and 70:30

IV. EXPERIMENTAL VALIDATION AND DISCUSSION

The experimental evaluation of the MDL-MEOLP method was performed using Real-time Iron Ore Feed Load Estimation [27]. This database (2.05GB/2.7GB uncompressed) consists of iron ore feed images with load classes (17 loads, tons). All images have dissimilar unknown ore categories. The only identified element is the ore load with respect to the image. Table II shows that the dataset consists of 17 classes with a total of 2,985 samples. Fig. 3 signifies the representative feed load images obtained from real-time mining operation dataset. Fig. 4 denotes the samples of preprocessed image. Fig. 5 represents the visual representation of learned feature maps.

The Real-time Iron Ore Feed Load Estimation dataset has a total of 2,985 images allocated across 17 feed load classes, with sample counts starting from 127 to 228. All images given were resized to 224 × 224 pixels for the model input. The dataset was collected from "Validation dataset for a new weakly supervised learning approach for Real-time Iron Ore Feed Load Estimation", to represent varying feed load levels. The dataset was partitioned into training and test sets using a stratified 70:30 split, ensuring proportional class representation in both subsets.

TABLE II. DETAILS OF DATASET

Classes	Number of Samples
Load 1	175
Load 2	207
Load 3	181
Load 4	166
Load 5	228
Load 6	169
Load 7	197
Load 8	170
Load 9	127
Load 10	211
Load 11	147
Load 12	184
Load 13	176
Load 14	164
Load 15	189
Load 16	152
Load 17	142
Total Samples	**2985**

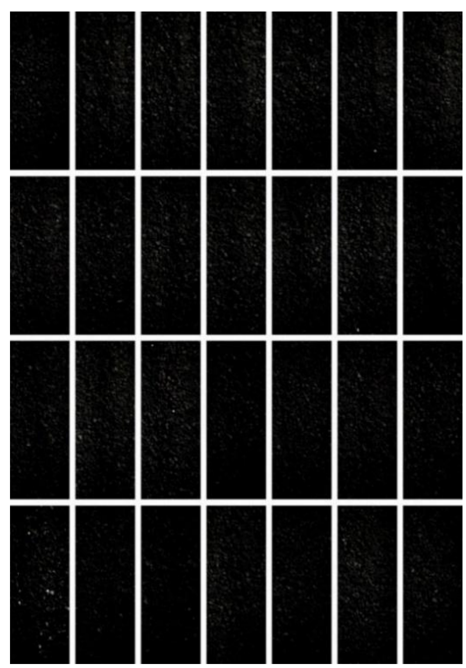

Fig. 3. Representative feed load images obtained from real-time mining operation dataset.

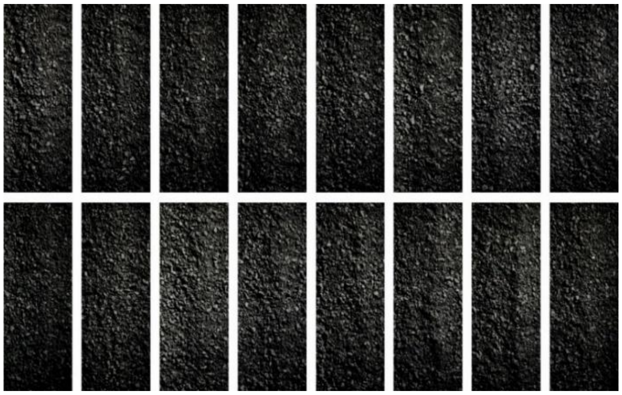

Fig. 4. Sample preprocessed images.

Fig. 6 displays confusion matrices for the diverse class feed load classifier method under 80:20 and 70:30 data splits. The diagonal dominance in each matrix suggests that most instances are appropriately categorized into each class. Some of the diagonal performance appear, presenting slight misclassifications among the similar categories. The outcomes confirm that the method preserves stronger and more reliable classifier values through data split.

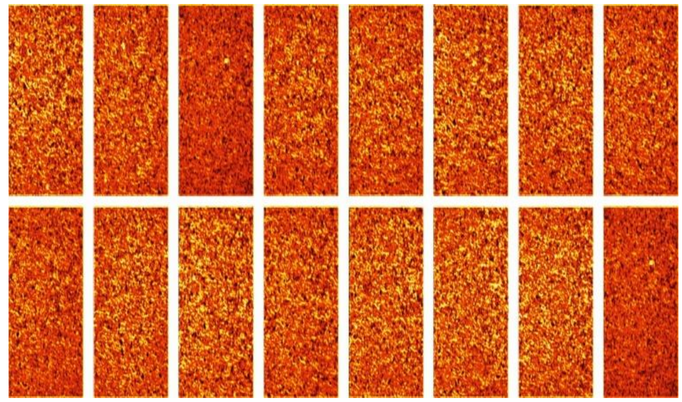

Fig. 5. Visual representation of learned feature maps.

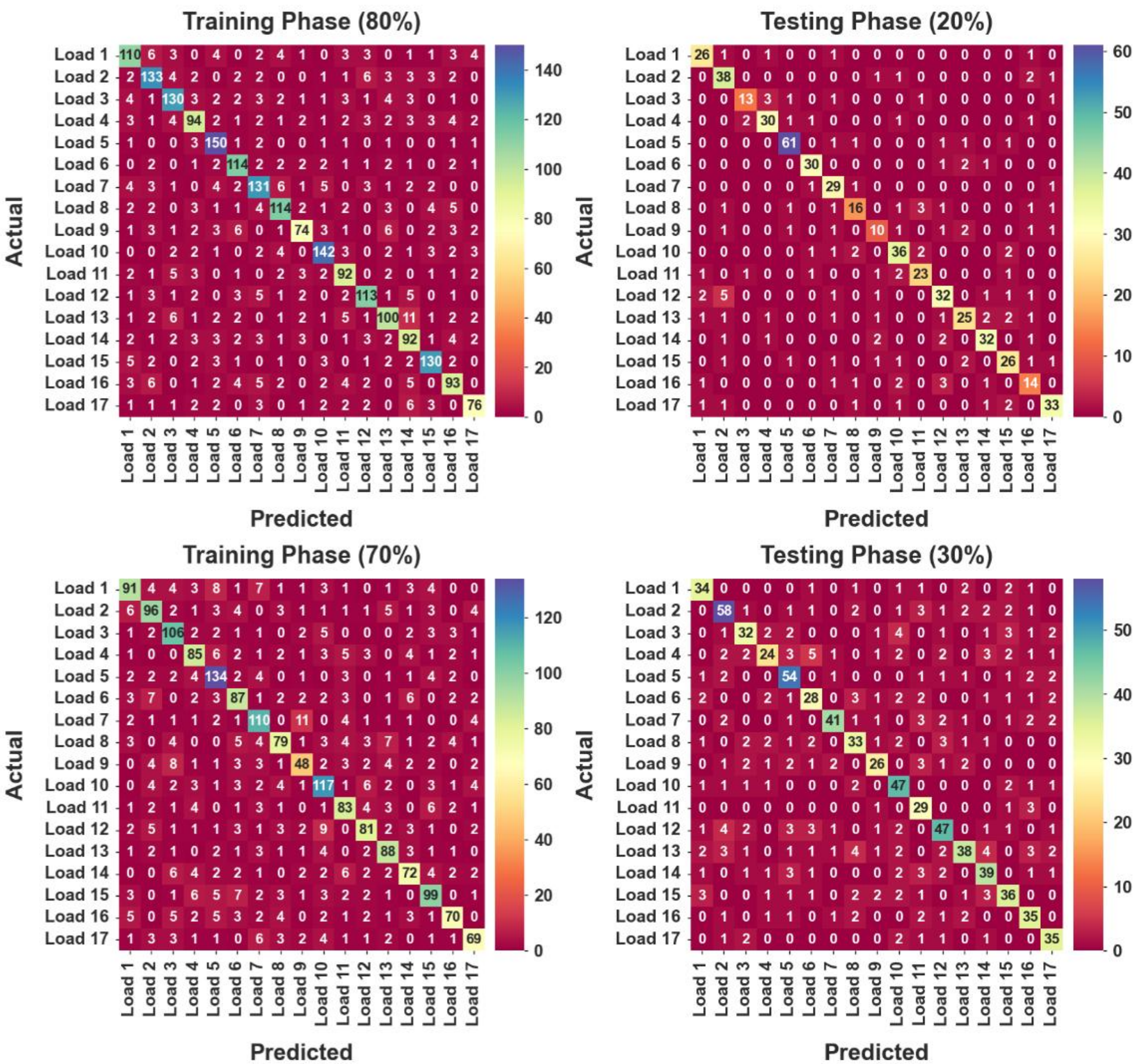

Fig. 6. Confusion matrix of the MDL-MEOLP method with 80:20 and 70:30 split.

Table III and Fig. 7 show the feed load classifier outcome of the MDL-MEOLP approach with an 80:20 train-test split. The result implied that the MDL-MEOLP method accurately detected the samples. The proposed MDL-MEOLP method attained an average value with *an accuracy* of 97.54% in TRAINING SET and *an accuracy* of 97.58% in TESTING SET, whereas the MCC of 77.30% for the TRAINING SET and MCC of 76.01% for the TESTING SET further confirm the model's reliable and balanced classification performance.

TABLE III. FEED LOAD CLASSIFIER RESULT OF THE MDL-MEOLP METHOD WITH 80:20.

Class	TRAINING SET (80%)	TESTING SET (20%)
Accuracy	97.54	97.58
Precision	78.77	77.75
Recall	78.50	77.01
F1-Score	78.57	77.19
MCC	77.30	76.01

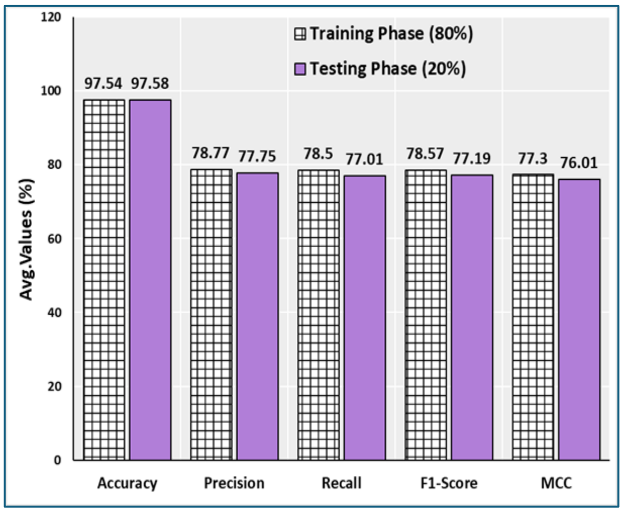

Fig. 7. Average value of the MDL-MEOLP method with 80:20.

Table IV and Fig. 8 exhibit the feed load classifier result of the MDL-MEOLP system with a 70:30 train-test split. The findings shown that the MDL-MEOLP approach precisely classified the instances. The proposed MDL-MEOLP technique reached average result with *an accuracy* of 96.77% in TRAINING SET and *an accuracy* of 96.59% in TESTING SET, while the MCC of 70.29% for the TRAINING SET and MCC of 68.88% for the TESTING SET more strengthen the model's consistent and stable classifier solutions.

Fig. 9 demonstrates the training and validation accuracy and loss graphs on 80:20 and 70:30. It shows the proposed model reaches increasing values over increasing epochs. In addition, the increasing validation over training exhibits that it learns efficiently on the test dataset. It also illustrates the loss investigation of the model. The results indicate that it reaches closer values of training and validation loss. It is observed that the MDL-MEOLP technique learns efficiently on the test dataset.

TABLE IV. FEED LOAD CLASSIFIER RESULT OF THE MDL-MEOLP METHOD WITH 70:30.

Class	TRAINING SET (70%)	TESTING SET (30%)
Accuracy	96.77	96.59
Precision	72.29	70.76
Recall	71.82	71.03
F1-Score	71.94	70.43
MCC	70.29	68.88

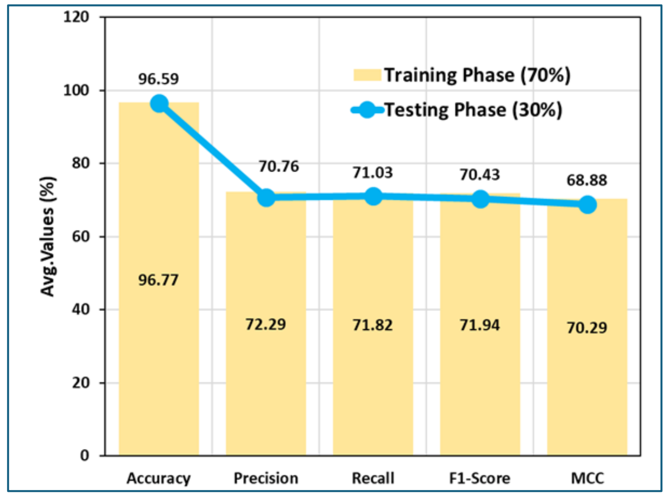

Fig. 8. Average value of the MDL-MEOLP approaches 70:30.

Table V and Fig. 10 portray the comparison study of the MDL-MEOLP approach with other methodologies using evaluation metrics [28, 29]. The stimulation outcome indicated that the MDL-MEOLP method achieves superior performance. The Xception, InceptionV3, ResNet50, DenseNet121 approaches have attained lower performance, whereas the EfficientNet-B0, MobileNetV3, ShuffleNetV2, and AlexNet methods obtain moderate performance. In addition, the EDF-NSDE reaches better prediction ability with more balanced outcomes. However, the proposed MDL-MEOLP method has achieved better overall solutions with *an acc*uracy of 97.58%, *prec*ision of 77.75%, *reca*ll of 77.01%, and an F1-score of 77.19%, representing its superior effectiveness and balanced classification capability compared to the other models. Furthermore, the reported values are directly referred to in the previously published study for a better comparison.

The gap that exists between accuracy (97.58%) and macro F1-score (77.19%) reflects the class imbalanced nature of the dataset, as macro averaging weights all classes equally and it is thus highly sensitive to the minority-class errors. It reports the confusion matrix and per-class metrics, showing that the misclassifications concentrate among visually similar and underrepresented adjacent load classes, which provides a more reliability beyond the accuracy alone.

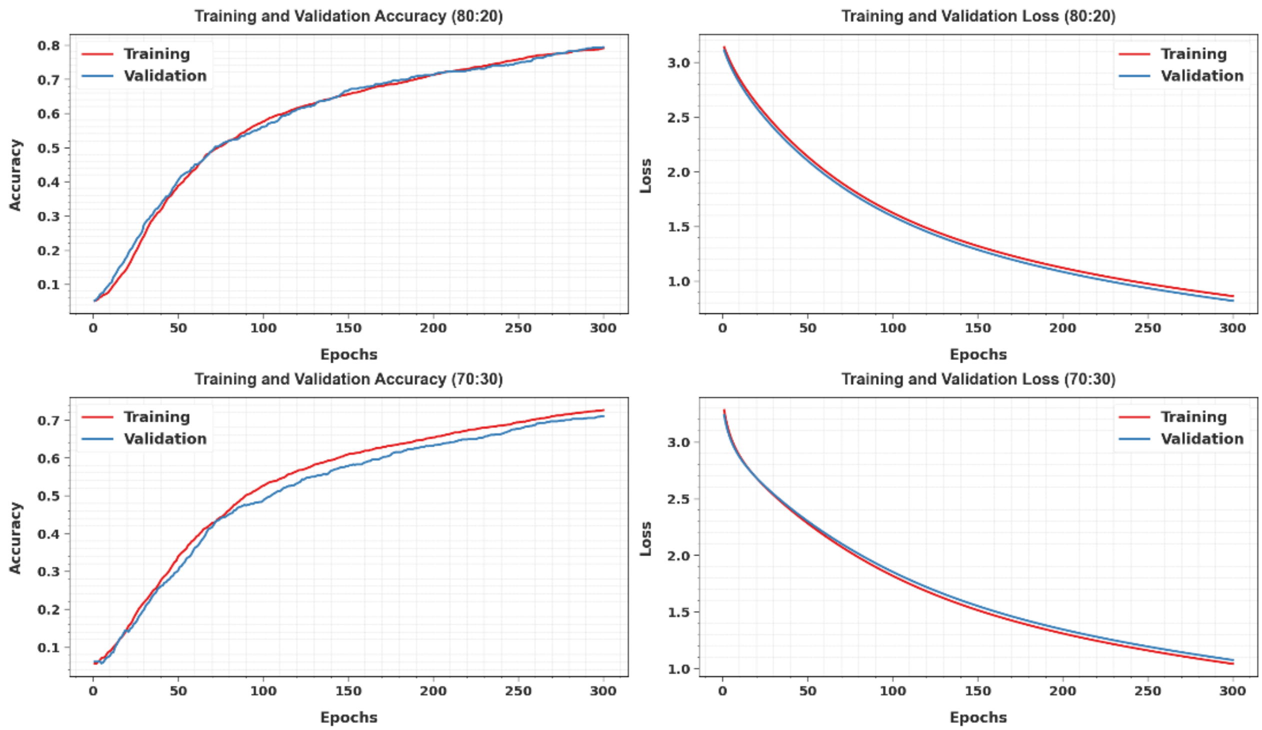

Fig. 9. Training and Validation Accuracy and Loss curve on 80:20 and 70:30.

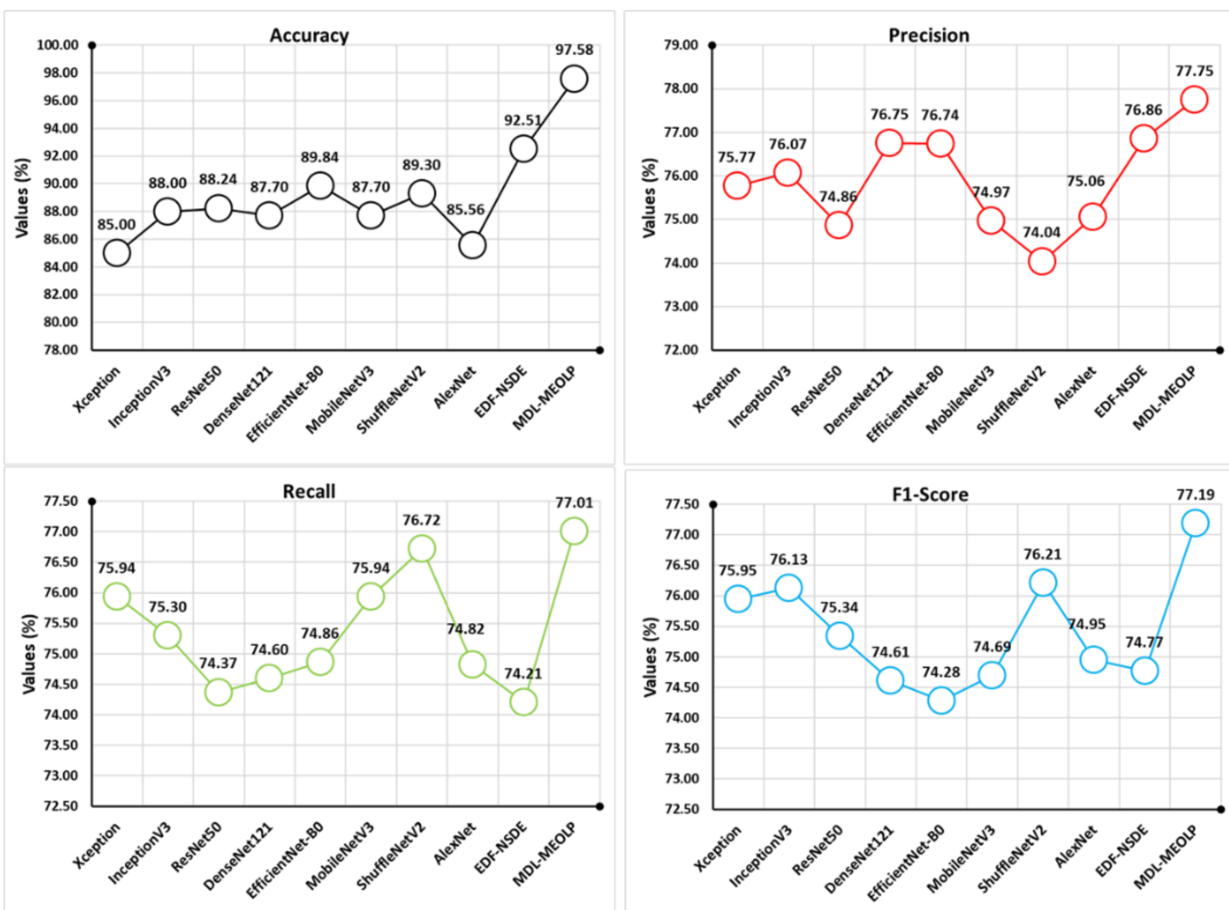

Fig. 10. Comparison outcome of the MDL-MEOLP approach with various metrics.

TABLE V. COMPARATIVE RESULTS OF THE MDL-MEOLP METHOD WITH EXISTING APPROACHES

Network	$Accur_y$	$Preci_n$	$Recal_l$	$F1_{score}$	Computational Time (sec)
Xception	85.00	75.77	75.94	75.95	13.85
InceptionV3	88.00	76.07	75.30	76.13	7.06
ResNet50	88.24	74.86	74.37	75.34	8.33
DenseNet121	87.70	76.75	74.60	74.61	9.12
EfficientNet-B0	89.84	76.74	74.86	74.28	10.13
MobileNetV3	87.70	74.97	75.94	74.69	10.94
ShuffleNetV2	89.30	74.04	76.72	76.21	8.07
AlexNet	85.56	75.06	74.82	74.95	8.12
EDF-NSDE	92.51	76.86	74.21	74.77	8.89
MDL-MEOLP	97.58	77.75	77.01	77.19	3.31

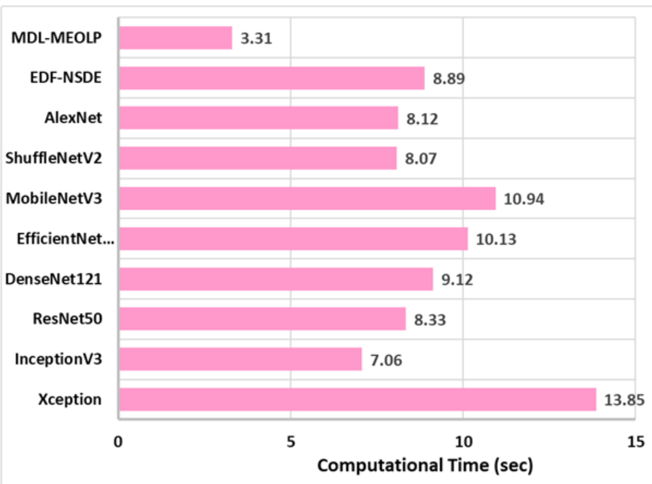

Fig. 11. Computational time of the MDL-MEOLP approach with other methodologies.

Fig. 11 reveal the computational time (CT) of the MDL-MEOLP approach with existing techniques. The result implied that the MDL-MEOLP method has gained minimal value with CT of 3.31sec, whereas the Xception, InceptionV3, ResNet50, DenseNet121, EfficientNet-B0, MobileNetV3, ShuffleNetV2, AlexNet, and EDF-NSDE require significantly higher processing time. This demonstrates that the MDL-MEOLP approach is faster and more computationally efficient than the competing methods.

An ablation study is used to analyze a model by systematically removing or altering its components to evaluate their effect on performance. It helps determine the importance of each module in the overall architecture. By comparing results, it reveals which parts contribute most to accuracy and efficiency. Table VI present the ablation study of the MDL-MEOLP method with metrics. The outcome implied that without Hierarchical Region Generation, EfficientNet-B0 only, EfficientNet-B0 + RDAM, Without Region-Density Attention Module, and Without Hierarchical Feature Fusion approaches achieve inferior value with various metrics, while the MDL-MEOLP (Region Generation + EfficientNet-B0 + RDAM + Hierarchical Fusion + Softmax) method obtain superior value with *an accu*racy of 97.58%, *prec*ision of 77.75%, *reca*ll of 77.01%, and F1-score of 77.19%, demonstrating the effectiveness of integrating all components.

The proposed MDL-MEOLP technique functions well for iron ore feed load forecast and mineral exploration applications. Enhancing image quality and extracting significant local details is supported by image preprocessing and grid-based patch extraction. The EfficientNet-B0 backbone efficiently learns density features and ore texture. The attention component assists the method, concentrated on critical ore areas, while minimizing background noise. For a better understanding of global and local data by combining hierarchical feature fusion. Finally, the approach attains higher precision in categorizing 17 feed load classes in comparison with other methodologies.

Table VII presents the k-fold cross-validation outcomes of the proposed MDL-MEOLP method. The method attained accuracy values ranged from 89.92% to 98.83%, with the maximum performance accuracy of 98.83% in K-6. The better precision of 98.43%, recall of 98.12%, and F1-score of 98.47% were achieved in K-6. Although slight difference over the folds, the constantly maximum evaluation measures illustrate the robustness, stability, and strong generalization ability of the proposed method.

TABLE VI. ABLATION STUDY OF THE MDL-MEOLP SYSTEM WITH METRICS

Network	Accuracy	precision	Recall	F1-Score
Without Hierarchical Region Generation	94.71	74.69	73.80	73.87
EfficientNet-B0 only	95.30	75.40	74.36	74.37
EfficientNet-B0 + RDAM	95.90	76.00	75.10	75.13
Without Region-Density Attention Module	96.41	76.57	75.78	75.70
Without Hierarchical Feature Fusion	97.04	77.17	76.43	76.44
MDL-MEOLP (Region Generation + EfficientNet-B0 + RDAM + Hierarchical Fusion + Softmax)	97.58	77.75	77.01	77.19

TABLE VII. K-FOLD CROSS VALIDATION RESULTS OF THE MDL-MEOLP SYSTEM.

K-Fold	Accuracy	Precision	Recall	F1-score
k-1	94.26	93.71	94.08	93.89
k-2	96.84	96.29	95.94	96.11
k-3	91.73	91.08	90.82	90.95
k-4	97.35	96.81	97.12	96.96
k-5	93.14	92.58	92.91	92.74
k-6	98.83	98.43	98.12	98.47
k-7	95.68	95.11	94.86	94.98
k-8	89.92	89.37	89.08	89.22
k-9	97.81	97.26	97.04	97.15
k-10	95.17	94.66	94.39	94.52

V. CONCLUSION AND FUTURE WORK

This study has presented an MDL-MEOLP approach for early-stage mineral exploration using iron ore images, which facilitates precise mining decision-making. Initially, the proposed model preprocesses the images to improve feature consistency and image quality. Besides, the hierarchical region generation is carried out using grid-based patch extraction to acquire multi-scale local regions. Moreover, an EfficientNet-B0 backbone is utilized for local feature extraction, enabling effective learning of texture patterns and ore density distributions. Furthermore, a region-density attention module is designed to emphasize dense ore regions while suppressing less informative areas through attention-based feature refinement. The hierarchical feature fusion is further applied to aggregate local-to-global representations and enhance global contextual understanding. Eventually, a classification head with GAP, dense layer, a dropout layer, and softmax activation is utilized to predict 17 feed load classes, and the model is trained using the Adam optimizer with cross-entropy loss over stratified data splits. To exhibit the improved performance of the proposed MDL-MEOLP model, a comprehensive experimental analysis takes place using Real-time Iron Ore Feed Load Estimation. The comparative results reported the effective performance of the MDL-MEOLP model with an accuracy of 97.58% over existing methodologies. Although the proposed algorithm functions well, it is primarily limited by its reliance on image-based information, which may fail to capture all geological and operational variations in mining settings. Furthermore, its high computational complexity may limit its viability in real-world applications.

Regardless the promising performance of MDL-MEOLP method, certain limitations remains. First, the dataset used is of limited in size and extracted from a single source, which may affects the generalization of ore feeds with various imaging conditions. It has not been verified on other ore types with substantially different density and texture distributions. The future work will be focussed on validating MDL-MEOLP model over larger and diverse multi-site datasets, which explores lightweight frameworks for real time edge deployments, and extending the architecture to related mineral exploration works such as ore-grade estimation and multi mineral classifications. Lastly, explainable AI methodology can be added to make the performance simpler to interpret for mining engineers.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://data.mendeley.com/datasets/b9ng2y2j5p/1 [27].

FUNDING STATEMENT

This research is funded by the Northern Border University through grant number (NBU-CRP-2026-249).

ACKNOWLEDGMENT

The authors extend their appreciation to the Northern Border University, Saudi Arabia, for supporting this work through project number (NBU-CRP-2026-249).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

[1] Saavedra, M., Risso, N., Momayez, M., Nunes, R. and Tenorio, V., 2026. Cluster-Based Machine Learning Modeling for Particle Size Variability in SAG Mill Feed. Mining, Metallurgy & Exploration, pp.1-14.

[2] Najafabadipour, A., 2026. Supervised Machine Learning-Based Determination of Groundwater Flow Direction in Complex Mining Areas. Mine Water and the Environment, pp.1-18.

[3] Yi, L., Yu, D., Wang, Y., Luo, B. and Peng, X., 2026. Ultra-short-term Forecasting Study of Power Load in Mega Steel Industry Based on Multi-stage Modeling. Recent Patents on Mechanical Engineering, 19(1), pp.131-149.

[4] Moyo, N., Mamvura, T., Danha, G. and Raghupatruni, P., 2025. Modelling and optimization of ultra-fine copper sulphide grinding: A hybrid statistical and machine learning approach. Particuology.

[5] Sil, J., Bhaumik, A.K. and Sen, S., 2025. Intelligent Real-Time Control Systems to Predict the Particle Size of Minerals Using Acoustics of Ball Mill (Vol. 600). Springer Nature.

[6] Liu, Y., Yan, G. and Pang, Y., 2025, May. Predictive Modeling of Ball Mill Load Parameters Using Hybrid Physics-Informed Neural Networks. In 2025 37th Chinese Control and Decision Conference (CCDC) (pp. 1604-1609). IEEE.

[7] Lucay, F.A., 2025. Balancing the Generation of Ultrafine, Fine and Coarse Particles in Comminution. Mineral Processing and Extractive Metallurgy Review, pp.1-23.

[8] Fang, Y., Ni, X., Xiao, Q., Huang, S. and López-Valdivieso, A., 2025. Iron-based materials synthesized by mechanical ball milling for environmental contaminants removal: progress and prospects. International Journal of Environmental Research, 19(1), p.12.

[9] Tripathi, P., Sunal, S., Dadwal, M., Kumar, D.S. and Ghorui, P.K., 2025. Development of an Automated Property Prediction System for the Pelletizing Process. Engineering Science & Technology, pp.247-260.

[10] Taghavi, J. and Gharabaghi, M., 2025. Prediction of ball mill power in iron ore concentration plants: A comparison between radial basis functions and linear regression. Results in Engineering, 26, p.105114.

[11] Parvathareddy, S., Yahya, A., Amuhaya, L., Samikannu, R. and Suglo, R.S., 2025. Transforming mining energy optimization: a review of machine learning techniques and challenges. Frontiers in Energy Research, 13, p.1569716.

[12] Behera, N., Laha, D. and Lahiri, S.K., 2025. Hybrid machine learning and genetic algorithm framework for optimising slag chemistry in electric arc furnaces. Canadian Metallurgical Quarterly, pp.1-17.

[13] Guo, H., Yang, Q., Yu, Y., Wang, L., Jia, P. and Pedrycz, W., 2026. Prediction of iron ore inventory at ports: A decomposition-integration hybrid approach incorporating key influencing factors. Expert Systems with Applications, 299, p.130041.

[14] Behera, N., Ghosh, A., Pal, P., Das, B. and Lahiri, S.K., 2025. Enhancing steelmaking productivity: a hybrid approach of machine learning and genetic algorithm optimisation for electric arc furnace. Ironmaking & Steelmaking, p.03019233251318881.

[15] Diniz, D.R.M., de Oliveira Morais, G., Reis, A.J.D.R. and Rêgo Segundo, A.K., 2026. A Soft Sensor for Inferring Energy Efficiency of Wet-Closed Ball Mills. Journal of Control, Automation and Electrical Systems, pp.1-18.

[16] Matos, S.N., Pinto, T.V., Duarte, R., Albuquerque, K.S., Fonseca, A.G., Ranieri, C.M., Marcolino, L.S., Pessin, G. and Ueyama, J., 2026. Data-driven soft sensor development for ore type estimation in mineral crushing processes. Engineering Applications of Artificial Intelligence, 167, p.113755.

[17] Najafabadipour, A., Hassanzadeh, F. and Kordestani, M., 2025. Advanced deep learning models for predicting elemental concentrations in iron ore mine using XRF data: a cost-effective alternative to ICP-MS methods. Environmental Geochemistry and Health, 47(4), p.104.

[18] Firdaus, S., Anwar, S. and Mohapatra, S., 2025. A deep hybrid inception network model with entropy-based attention for automated iron ore image characterization. Scientific Reports, 15(1), p.38567.

[19] Giannetti, C., Borghini, E., Carr, A., Raleigh, J. and Rackham, B., 2025. Deep learning for robust forecasting of hot metal silicon content in a blast furnace. The International Journal of Advanced Manufacturing Technology, 136(9), pp.4045-4054.

[20] Park, S., Jung, D. and Choi, Y., 2023. Prediction of ore production in a limestone underground mine by combining machine learning and discrete event simulation techniques. Minerals, 13(6), p.830.

[21] Zou, G., Zhou, J., Song, T., Yang, J. and Li, K., 2023. Hierarchical intelligent control method for mineral particle size based on machine learning. Minerals, 13(9), p.1143.

[22] Luo, X., Huang, W., He, S., Xiao, W. and Jiang, Z., 2025. Deep Feature Decoupling Network for Ball Mill Load Signals. Machines, 13(10), p.881.

[23] Polatgil, M., 2022. Investigation of the effect of normalization methods on ANFIS success: forestfire and diabets datasets. International Journal of Information Technology and Computer Science, 14(1), pp.1-8.

[24] Hussain, A., Ul Amin, S. and Fayaz, M., 2023. An Efficient and Robust Hand Gesture Recognition System of Sign Language Employing Finetuned Inception-V3 and Efficientnet-B0 Network. Computer Systems Science & Engineering, 46(3).

[25] Zhou, C., He, Z., Dong, J., Li, Y., Ren, J. and Plaza, A., 2025. Low-rank and sparse representation meet deep unfolding: A new interpretable network for hyperspectral change detection. IEEE Transactions on Geoscience and Remote Sensing.

[26] Ganguly, P. and Ghosh, A., 2024, November. Efficient brain tumor classification with lightweight CNN architecture: a novel approach. In 2024 IEEE International Conference on Future Machine Learning and Data Science (FMLDS) (pp. 325-330). IEEE.

[27] L. Guo, "Validation dataset for a new weakly supervised learning approach for real-time iron ore feed load estimation," Data Archiving and Networked Services (DANS). Apr. 29, 2022. doi: 10.17632/b9ng2y2j5p.1. Available at: https://data.mendeley.com/datasets/b9ng2y2j5p/1

[28] Firdaus, S., Anwar, S. and Mohapatra, S., 2025. A deep hybrid inception network model with entropy-based attention for automated iron ore image characterization. Scientific Reports, 15(1), p.38567.

[29] Zhang, D., Qian, X., Shi, C., Zhang, Y., Qian, Y. and Zhou, S., 2025. Iron Ore Image Recognition Through Multi-View Evolutionary Deep Fusion Method. Future Internet, 17(12), p.553.