SHAP Guided Transformer-Based Deep Learning Model for Diamond Clarity Grading in Smart Mining and Resource Valuation Systems

Ahlem Fatnassi[1], Shouki A. Ebad[2]*, Sami F Karali[3], Mohammed I Thanoon[4],
Majed Nawas[5], Monir Abdullah[6], Asma A. Alhashmi[7], Abdulbasit A. Darem[8]
Department of Computer Science-College of Science, Northern Border University, Arar 73213, Saudi Arabia[1, 5, 7]
Center for Scientific Research and Entrepreneurship, Northern Border University, 73213, Arar, Saudi Arabia[2, 8]
Computing Department-College of Engineering and Computing at Al-Lith, Umm Al-Qura University, Saudi Arabia[3, 4]
Department of Computer Science and Artificial Intelligence-College of Computing and Information Technology,
University of Bisha, Bisha 61922, Saudi Arabia[6]

***Abstract*—Diamond clarity grading aims to automatically determine the quality of diamonds by analyzing visible inclusions and structural flaws using intelligent analysis methods. By integrating diamond cost prediction with clarity grading, a unified architecture can be employed for both quality examination and market value evaluation, where clarity attributes support the determination of the diamond's economic value. Traditional machine learning (ML) algorithms depend upon handcrafted characteristics that frequently miss composite structural and visual patterns. Deep learning (DL) enhances feature learning; however, it faces difficulties such as data imbalance, limited interpretability, and inadequate generalization. Consequently, the real-time diamond grading outcome remains limited. Therefore, this study proposes a SHAP-Guided Transformer Network (SGT-Net) for diamond clarity grading. The proposed model includes a preprocessing stage that prepares the diamond grading dataset for effective and reliable model training. Moreover, feature engineering is carried out to enhance the representation of the dataset by deriving meaningful attributes from the raw diamond features. Moreover, SHAP-guided feature prioritization is applied to identify the most influential features in the data. The transformer-based architecture processes the data using separate embeddings for numerical and categorical features, which are then passed through a transformer encoder with multi-head self-attention. In this setup, SHAP is used to understand feature importance and guide the model to pay more attention to the most relevant features during learning. The resulting fused feature representations are then passed through a dense layer followed by a softmax classifier to predict diamond clarity into 8 distinct classes. For training, the model is optimized using the RAdam optimizer along with a weighted cross-entropy loss function. Experimental evaluation is conducted using the Diamond dataset with 8 clarity classes and 53940 samples from the Kaggle repository. The results demonstrate that the proposed SGT-Net effectively learns complex feature relationships and improves diamond clarity grading performance with enhanced accuracy of 96.59%, making it suitable for resource valuation systems.**

Keywords*—Diamond clarity grading; quality assessment; deep learning; multi-head self-attention; model optimization

*Corresponding author.

I. INTRODUCTION

Diamond is the hardest natural material and is broadly utilized in industrial fields because of its distinctive physical characteristics. Since the 1960s, synthetic diamond technology has developed and matured, and it has completely substituted natural ones in industrial applications [1]. The unique physical characteristic of diamonds is closely associated with their crystal structure, which is considered a unique consumer product. Diamond is the hardest mineral, about 58 times harder than other types of minerals, and is utilized in several machines and other types of tools for cutting and slicing processes [2]. It is also one of the most expensive gemstones among the other gemstones; its popularity is mainly due to its unique optical characteristics. Additional factors include its durability, fashion, custom, and strong marketing by diamond producers [3]. However, quality assessments are still highly based on manual inspection through optical microscopes; this method of inspection is time-consuming and prone to operator fatigue and subjectivity, with limited efficiency, reliability, and quantitative evaluation. With the growing demand for accuracy and automated systems, automated visual inspection is becoming an essential trend in industrial diamond processing [4].

A major factor that affects the rate of the diamond is its clarity, which is primarily assessed based on the number, type, position, and size of the internal inclusions [5]. It is a very costly and time-consuming process, and the result may vary depending on the operator's subjective experience. Achieving the accuracy and reproducibility of the outcomes requires proper training and verification by multiple experts, and standardized procedures are essential [6]. Over the past few years, Deep Learning (DL) models have delivered an innovative pathway for the automatic recognition of diamond inclusions. However, the current approaches mostly depend on target positioning and do not entirely incorporate the key features of diamond clarity [7]. These methods overlook the interpretability transformation from detection to grading, which makes the output of the model less suitable for industrial adoption.

Especially, based on the two limitations of low accuracy in micro-inclusion detection and unclear logic for defining clarity grades, developing an automatic grading system with higher precision and consistency remains a difficult task [8].

Since the development of AlexNet in 2012, convolutional neural networks (CNNs) have played a vital role in the rapid development of computer vision. Object detection techniques, including YOLO, Faster R-CNN5, and instance segmentation methods like Mask R-CNN, and a point-based mask refinement approach, such as PointRend, achieve effective object localization and feature extraction by capturing channel and spatial representations in local regions [9]. In recent times, lightweight CNN detectors are used for synthetic diamond quality analysis, with enriched YOLOv8n-driven approaches that have proven effective in capturing local morphological features in microscopic imaging systems [10]. Although anchor-based one-stage detectors are efficient for real-time applications, they are primarily focused on local feature aggregation and scale-based fusion. Alternatively, the multi-head self-attention (MHSA) module in transformers is highly effective for capturing long-term dependencies and less capable of capturing local features [11].

Although diamonds are very highly valued for gemstone and industrial applications, their clarity grading remains largely accomplished by manual microscopic inspection, which is slower and more subjective. Conventional DL methodology mainly concentrates on identifying inclusions instead of directly mapping them to clarity grading. CNN-based strategy is efficient in extracting local attributes; however, it struggles to model global dependencies, whereas Transformers extract global context but are weak in delicate feature representation. Therefore, there is a requirement for an automatic and consistent methodology that combines both local and global feature learning for precise diamond clarity grades. To address these limitations, this research presents a SHAP-Guided Transformer Network (SGT-Net) for diamond clarity grading. The major contributions are outlined below:

- Construct engineered features like diamond volume, carat-to-volume ratio, and depth percentage to enhance feature representation and capture intrinsic features.
- Utilize SHAP-based feature prioritization to identify and emphasize the most influential features, enabling the model to focus on high-impact attributes during learning.
- Design a Transformer-based model with a feature tokenizer (FT) that separately embeds numerical and categorical features, followed by multi-head self-attention (MHSA) with SHAP-guided refinement to capture important feature interactions and improve learning efficiency, and a feature fusion layer to integrate the learned representations.
- Apply a dense layer followed by a softmax classifier to predict diamond clarity into eight distinct classes.
- Integrate the RAdam optimizer along with a weighted cross-entropy loss function to effectively optimize the model during training.
- Conduct the experimental outcomes using numerous metrics, and the comparison analysis assures that the proposed SGT-Net model gains promising results compared to existing models.

II. Related Work

This section presents a review of existing studies on diamond clarity grading and highlights the identified research gaps. Yan et al. [12] developed a target recognition technique specially created for quality estimation of synthetic diamonds, called Diamond-DETR, targeting to enhance identification precision and method generalization across constrained resource settings. The proposed model augments the backbone model to introduce a multi-scale feature representation and employs the RepFasterNet module to maximize inference speed and minimize computational complexity. Moreover, it combines an encoder part that depends on a higher-range screening-feature fusion that applies the channel attention system to fuse and filter attributes with distinct measures, thus improving the method's capability to identify the shortcomings of multiple dimensions. Zhang et al. [13] designed an improved YOLOv8n strategy for synthetic diamond identification and quality calculation, known as YOLOv8n-adamas. The author reshaped the architecture to enrich feature extractor abilities and enhance model performance. The dynamic detection head depends on the attention mechanisms (AM). Mahmood and Zia [14] examined the crucial characteristics to estimate the hardness of DLC coatings. To expand the small experimental data of DLC coatings, the conditional tabular generative adversarial networks (CTGANs) method has been applied. For predicting the hardness of DLCs, a range of 15 ML methodologies are utilized, and their efficiency is evaluated using 6 common error-based evaluation methods.

Triet et al. [15] investigated the incorporation of blockchain technology to overcome these limitations, exploiting the abilities of Fantom and BNB Smart Chain. To collect and signify vital data about each diamond, Non-Fungible Tokens (NFTs) are employed; our system improves traceability and transparency. Through the generation of the protective, immutable data of every diamond, blockchain technology can substantially enhance diamond quality management. In [16], a defective segmentation pipeline is presented by employing in-situ optical images to detect attributes that specify imperfect conditions, which are observed at the macro-scale. In this previous work, a semantic segmentation algorithm was used; these imperfect states and the corresponding derived characteristics are separated and categorized by the pixel masks. Yang et al. [17] discussed how to eradicate the inconsistency and subjectivity inherent in the standard International Association of Drilling Contractors (IADC), which relies heavily on the expertise of drilling engineers and frequently causes unstable performance. Utilizing improvements in CV and DL approaches for polycrystalline diamond compact (PDC) bit damage, this study proposed an automatic identification strategy. For finding the PDC bit cutters, YOLOv10 was applied, accompanied by a dual SqueezeNet system to perform wear type and wear rating categorizations.

Basha et al. [18] addressed diamond rates by WEKA's data mining software. In this study, diamond data is effectively used. These algorithms include Random Forest, M5P, Decision Stump, Multilayer Perceptron, M5Rules, and REP Trees. To evaluate the price of a diamond, cut quality, carat weight, and how well they have accessible size, such as depth, breadth, and the table's surface, diverse ML strategies are contrasted and compared. Randhawa et al. [19] proposed the use of a regression approach in ML to predict diamond costs. The dataset involves attributes like color, carat, clarity, and cut, with certain physical sizes, namely x, y, z, table, and depth, that are both continuous and categorical. Several categorical characteristics can be encoded as constant values to simplify and improve prediction.

A. Research Gap

Despite recent developments, numerous gaps remain in existing studies. Diamond-DETR focuses primarily on detection and not on direct clarity grading, thus making it of low value for final grading decisions. The YOLOv8n-adamas enriches feature extraction but still depends on CNNs, which reduces its capability to capture global relations required for precise clarity valuation. The segmentation-based techniques decrease labeling effort but rely on massive numbers of annotated datasets, and they cannot directly convert the defects into clarity grades. Existing tabular Transformer models mainly learn relationships between features through self-attention, but they do not directly use feature importance to give greater focus to the most relevant attributes. Gradient-boosting methods can learn nonlinear relationships among the input features, but they have limited ability to capture the wider dependencies between them. Therefore, existing methods may not sufficiently capture both the importance of individual features and their interactions, which can affect the accuracy of clarity classification. Accordingly, there is still a requirement for a combined and effective system for accurate diamond clarity grading.

III. Materials and Methods

This study presents an SGT-Net approach for diamond clarity grading in smart mining. Fig. 1 represents the structural overview of the proposed SGT-Net approach. The proposed SGT-Net follows a multi-stage approach. First, data preprocessing is performed through median imputation, duplicate record removal, outlier detection, ordinal encoding for cut, color, and clarity, and feature standardization using RobustScaler. Then, the feature engineering process is performed to derive meaningful attributes from raw data. In addition, the SHAP-based features identify and rank the most influential features. The transformer-based architecture is processed using numerical and categorical feature embedding, transformer encoder, MHSA, SHAP-based refinement, and a feature fusion layer, followed by a dense classification head and softmax activation to predict 8 diamond clarity classes. Finally, the Adam optimizer with a weighted cross-entropy loss function is used to improve training stability.

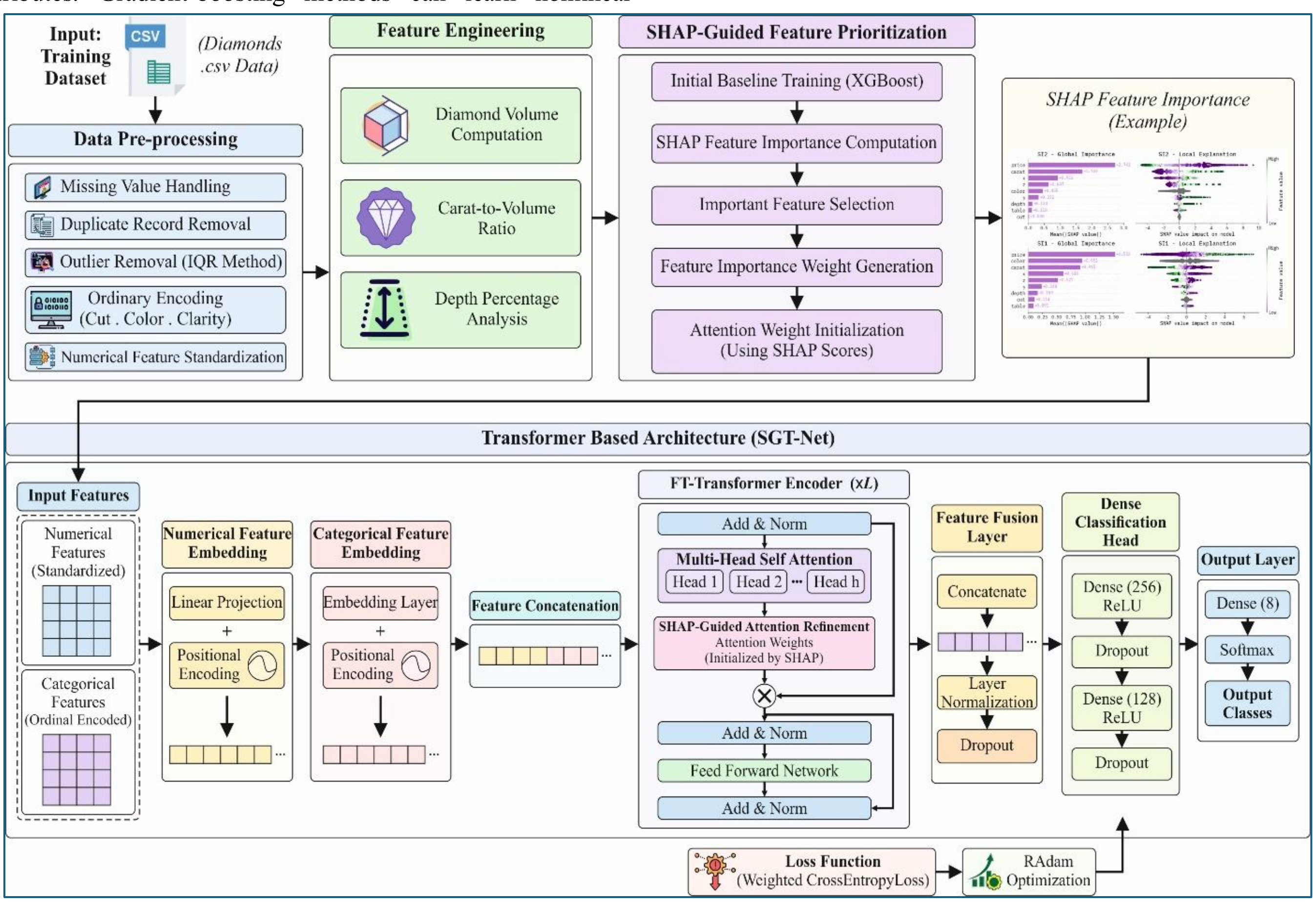

Fig. 1. Structural overview of the proposed SGT-Net model.

A. Dataset Used

The proposed SGT-Net framework is evaluated using the diamonds dataset [20]. The dataset includes information about diamonds and is mainly used to estimate their market value based on various physical and quality-based features. It contains both numerical and categorical features that describe details of each diamond. Quality-related features, including cut, color, and clarity, describe the visual characteristics and grading of a diamond. Physical measurements such as depth provide further information about its proportions. Together, these features provide useful information for building predictive models for diamond pricing and clarity assessment in smart mining and resource valuation applications.

Table I presents the distribution of 53,940 diamond samples across eight clarity classes. The clarity grades range from I1, representing the lowest grade, to IF, representing the highest grade. Among these classes, SI1 has the largest number of samples (13,065), followed by VS2 (12,258) and SI2 (9,194), providing good representation of diamonds with moderate clarity levels. Overall, the dataset covers all eight clarity grades and provides sufficient samples for training and evaluating the classification model.

TABLE I. DATA DISTRIBUTION OF DIAMOND CLARITY CLASSES

Clarity Classes	Description	No. of Samples
I1	Included 1 (worst clarity)	741
SI1	Slightly Included 2	13065
SI2	Slightly Included 1	9194
VS1	Very Slightly Included 2	8171
VS2	Very Slightly Included 1	12258
VVS1	Very Very Slightly Included 2	3655
VVS2	Very Very Slightly Included 1	5066
IF	Internally Flawless (best clarity)	1790
Total Samples		**53940**

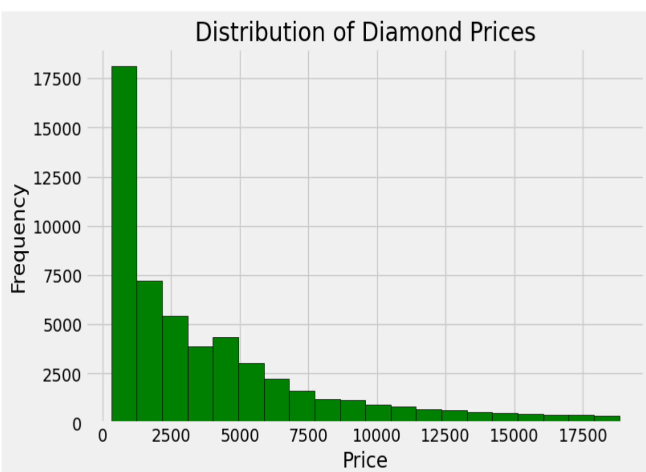

Fig. 2. Distribution of diamond price.

Fig. 2 presents the distribution of diamond prices in the dataset. The histogram shows that most diamonds have low to moderate prices, while only a small number of samples have higher prices, resulting in a right-skewed distribution. This pattern gives a clear view of the price variation in the dataset and provides a useful basis for further analysis.

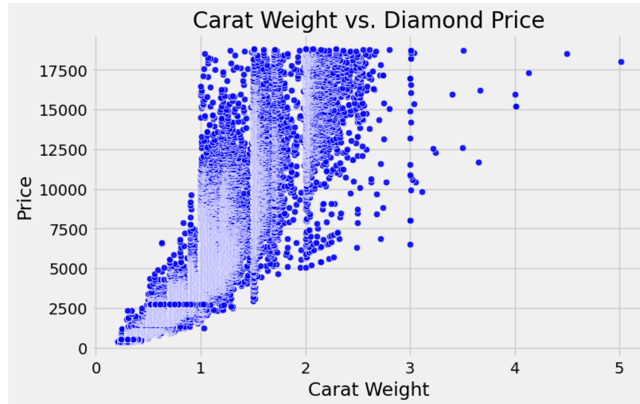

Fig. 3. Relationship between carat weight and diamond price.

Fig. 3 displays the relationship between the carat weight and diamond price. The scatter plot shows a strong positive relationship between carat weight and diamond price, indicating that the diamond price grows gradually with higher carat values. This visualization represents the importance of carat weight in defining overall pricing.

B. Data Preprocessing

Data preprocessing is a vital step that improves data quality and makes the dataset suitable for model training. It includes median imputation for missing values, removing duplicate records, detecting and handling outliers, encoding features like cut, color, and clarity, and scaling numerical features using RobustScaler.

1) Median imputation: Median imputation fills missing entries using the median value calculated from the corresponding feature. Median imputation replaces missing values with the median value of the respective attribute within the specified class [21].

$$\hat{x}_{ij} = median\left(x_{ij}: x_{ij} \in C_k\right) \quad (1)$$

The imputation method can be improved by using class-wise median values to better preserve the data distribution and reduce the influence of outliers.

2) Duplicate record removal: It is performed to confirm the quality of data and prevent repetition in the dataset. Recurrent occurrence of the same documents is detected and removed, thereby ensuring that every record contributes exclusively to the learning process. This stage helps avoid bias in model training and enhances the consistency of the dataset.

3) Outlier removal using IQR method: Outliers are eliminated by the Interquartile Range (IQR) strategy to minimize the effect of extreme values and enhance data consistency. The system recognizes anomalous data points that lie substantially outside the central distribution of every numerical attribute. These extreme values are eliminated to guarantee that the dataset replicates the true underlying patterns and improves method consistency.

4) Ordinal encoding: Ordinal encoding is employed for categorical characteristics, namely, color, clarity, and cut, by transferring them into numerical values that depend on their inherent order. Every category is allotted a rank based on its level of quality, maintaining the natural hierarchy presented in the information. This convention facilitates the method to effectively interpret categorical data in a structured numerical form.

5) Numerical feature standardization: RobustScaler is a feature scaling technique applied in data preprocessing that removes the median and scales numerical attributes based on the IQR. It is remarkably efficient for datasets including outliers, due to the fact that, unlike Min-Max Scaling or Standardization, it is not affected by extreme values.

$$\mathrm{x}' = \frac{x - Median(X)}{Q_3 - Q_1} \tag{2}$$

Here x refers to the scaled value, and x denotes the original feature value.

C. Feature Engineering

Feature engineering includes valuable features from raw diamond information to enhance method learning, including evaluating diamond volume (x × y × z), capturing the carat-to-volume ratio, examining depth proportion, and generating interaction-based attributes among volume, depth, and carat. These feature engineers are standardized to confirm effective and durable extraction for downstream modeling. Furthermore, these transformations help capture hidden geometric and physical relationships that are not expressed particularly in the original dataset. The increased feature space enhances the separability between several clarity types to provide the most discriminative patterns. Therefore, feature engineering empowers the method's capability to learn compound dependencies and enhance predictive outcomes.

D. SHAP-Guided Feature Prioritization

SHAP is a method used to explain the output of ML methodologies by measuring the contribution of each feature to a prediction. SHAP is used for interpreting the predictions produced by the XGBoost technique, which is frequently observed as a black box owing to its complexity [22]. SHAP values expose the extent to which every input attribute impacts the method's outcome, in both global and local stages. This system enhances the transparency of the algorithm, to verify predictions depend on their field knowledge, by allowing energy analysts. By using the Shapley value formula, the SHAP values are computed:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!\,(|N| - |S| - 1)!}{|N|!} [v(S \cup \{i\}) - v(S)] \tag{3}$$

here: S represents a subset of features not containing feature i, $v(S)$ signifies the value function demonstrating the model output with features in the subset S, and N indicates the set of all features. This expression has time complexity; the Tree SHAP methodology calculates SHAP values in a given time, where D denotes the highest tree depth. Additionally, SHAP offers reliable attributes, ensuring that characteristics contribute more to predicting the highest significance scores stably. This makes it specifically appropriate for an intricate ensemble approach, where interpretability is restricted. During practical applications, SHAP also supports detecting hidden interactions between attributes, which can further influence feature engineering and model refinement.

The SHAP values obtained from the XGBoost model are used to determine the importance of each input feature. Since both positive and negative SHAP values indicate the strength of a feature's contribution, their absolute values are considered for feature prioritization. The absolute SHAP values are normalized using min–max normalization as follows:

$$s_i = \frac{|\phi_i| - \min(|\phi\,|)}{\max(|\phi\,|) - \min(|\phi\,|) + \epsilon} \tag{4}$$

where, ϕ denotes the SHAP value of the i_{th} feature, $|\phi_{\,i}|$ represents the magnitude of its contribution to the prediction, and s_i is the normalized importance weight of that feature. The term ϵ is a small constant introduced to avoid division by zero. A higher value of s_i indicates a stronger contribution of the corresponding feature to the model prediction.

The normalized SHAP weights are then incorporated directly into the attention mechanism. The SHAP values do not modify the input feature embeddings or the query, key, and value representations.

$$Q = XW_Q, K = XW_K, V = XW_V \tag{5}$$

where, X represents the input feature representations, and W_Q, W_K, and W_V implies the learnable projection matrices for the query, key, and value representations. The scaled dot-product attention score is then modified by adding the normalized SHAP importance of the corresponding key feature:

$$A_{ij} = \frac{Q_i K_j^T}{\sqrt{d_k}} + \lambda s_j \tag{6}$$

where, Q_i and k_j are the query and key representations of the i_{th} and j_{th} features, d_k is the dimension of the key vector, s_j is the normalized SHAP importance of the j_{th} feature, and λ controls the contribution of the SHAP-based importance to the attention score. Consequently, the SHAP information is introduced only at the attention-logit level and does not alter Q, K, or V. The modified attention logits are normalized across the key dimension using the softmax function:

$$\alpha_{ij} = \frac{\exp(A_{ij})}{\sum_k \exp(A_{ik})} \tag{7}$$

Here α_{ij} signifies the attention probability assigned by the i_{th} query to the j_{th} key. The SHAP-guided attention output is subsequently calculated as:

$$O_i = \sum_j \alpha_{ij} V_j \tag{8}$$

where, O_i represents the output corresponding to the j_{th} query position, therefore, features with higher SHAP importance receive a stronger positive bias in the attention logits, resulting in greater attention probability when the other attention components are comparable.

This allows the Transformer to emphasize features that have a stronger contribution to the XGBoost prediction while preserving the learned feature representations and Q/K/V projections. The resulting attention representation is then forwarded to the subsequent Transformer layer or classification module. Algorithm 1 represents the pseudocode for SHAP-Guided Transformer Attention:

Algorithm 1: Pseudocode of SHAP-Guided Transformer Attention

Input: Feature matrix X, trained XGBoost model, Transformer parameters, SHAP weighting coefficient λ, and a small constant ϵ

Output: SHAP-guided attention representation O

1. Train the XGBoost model using the training feature matrix X.

2. Compute the SHAP values ϕ for the input features using the trained XGBoost model.

3. Calculate the absolute SHAP values:

$$|\ \phi_i\ |, i = 1,2, \dots, N$$

4. Normalize the absolute SHAP values using min-max normalization:

$$s_i = \frac{|\ \phi_i\ | - min(|\ \phi\ |)}{max(|\ \phi\ |) - min(|\ \phi\ |) + \epsilon}$$

5. Generate the query, key, and value representations:

$$Q = XWQ, K = XWK, V = XWV$$

6. Associate each normalized SHAP importance s_j with the corresponding j_{th} key feature.

7. Compute the SHAP-guided scaled dot-product attention logits:

$$A_{ij} = \frac{Q_i K_j^T}{\sqrt{d_k}} + \lambda s_j$$

8. Apply softmax across the key dimension to obtain the attention probabilities:

$$\alpha_{ij} = \frac{exp(A_{ij})}{\sum_k exp(A_{ik})}$$

9. Compute the SHAP-guided attention representation:

$$O_i = \sum_j \alpha_{ij} V_j$$

10. Pass the resulting representation O to the subsequent Transformer layer or classification module.

E. Transformer-Based Architecture

The proposed SGT-Net uses a transformer-based architecture that processes numerical and categorical features through separate embeddings and an FT-Transformer encoder with MHSA. SHAP-guided refinement and feature fusion are applied before passing the learned representations to a dense layer and softmax classifier for predicting 8 diamond clarity classes.

1) Feature tokenizer: This model is responsible for converting the input feature values into trainable embeddings [23]. Particularly, it captures two parts, such as numerical (x_i^{num}) and categorical (x_i^{cat}) variables, which are embedded in diverse methods. Here, W_i^{num} implies the weights of the numerical embedding layer, and b_i^{num} denotes the numerical bias term; embed () represents the categorical embedding operation, and b_i^{cat} implies the categorical bias term.

$$E_i^{num} = x_i^{num} \cdot W_i^{num} + b_i^{num} \tag{9}$$

$$E_i^{cat} = embed(x_i^{cat}) + b_i^{cat} \tag{10}$$

Also, the numerical and categorical embeddings and a trainable classifier embedding E^{cls}, that all feature embeddings have formed and integrated with each feature embedding.

*2) Transformer encoder***:** The transformer encoder module learns tabular feature representations by representing every feature as an individual token [24]. With the input feature vector $x = [x_1, x_2, \dots, x_7]^T$, all features are self-sufficiently mapped into a common embedding space utilizing a linear transformation:

$$e_j = x_j W_{lin} + b_{lin} \tag{11}$$

Here, $W_{lin} \in \mathbb{R}^{1\times d}$and $b_{lin} \in \mathbb{R}^{d}$. To maintain feature identity, a trainable positional embedding $p_j \in \mathbb{R}^{d}$ has been inserted:

$$z_j = e_j + p_j \tag{12}$$

These tokenized feature representations enable the model to capture the unique characteristics of each input attribute while maintaining a consistent embedding structure. The resulting token sequence is then forwarded to the transformer encoder, where relationships and dependencies among features are learned through the self-attention mechanism.

3) Multi-head self-attention: The attention mechanism (AM) was initially proposed in neural machine translation, improving the capability of the method to concentrate on essential data through weighted allocation [25]. The MHSA module is able to evaluate the effect of input features on the reconstruction curves and allocate corresponding weights to all features. It emphasizes data that significantly impacts the reconstruction curve, thus enhancing model precision. It trains the information's multi-layered interaction across numerous independent focal points and uses a structure with h heads.

$$Q_i = XW_{Qi}, K_i = XW_{Ki}, V_i = XW_{Vi} \tag{13}$$

Next, use the scaled dot-product attention score operation for all attention heads' Q, K, and V to compute the attention scores A:

$$A_i = soft\max\left(\frac{Q_i K_i^T}{\sqrt{d_k}}\right) V_i \tag{14}$$

In this Eq. (10), d_k implies the attention head dimension,· and the softmax function processes attention weights. Finally, the outcomes from all h attention heads are integrated and linearly projected utilizing the parameter matrix W_O to attain the resulting value:

$$\text{MultiHead}(X) = Concat(A_1, \dots, A_h) W_O \tag{15}$$

4) SHAP-guided attention refinement: SHAP is utilized for quantifying the contribution of all features, allowing interpretability-based attention score weights. It improves the transformer AM by highlighting higher-effect features and suppressing less relevant inputs.

5) Feature fusion layer: The feature fusion layer integrates diverse kinds of learning representations, like numerical and categorical embeddings, and attention-refined attributes, into a single combined feature vector. It supports the method of combining complementary data from numerous sources and high feature relationships for enhanced predictive outcomes.

6) Dense classification head: It is applied to transform the fused feature representation into class-by-class predictive scores. It serves as the final learning layer in the system, allowing decision-making to rely on deep feature extraction. The fused feature vector is mapped to class logits employing a fully connected layer:

$$z = Wx + b \tag{16}$$

Here, x denotes the fused feature representation, W implies the matrix of weights, and b denotes the bias terms.

7) Softmax output: The softmax function transforms logits into class probabilities through eight diamond clarity grades, assuring a standardized probabilistic outcome appropriate for multi-class identification.

$$Softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_i}} \tag{17}$$

Let z_i imply the logit for class i, and K signifies the overall count of classes. Fig. 4 signifies the transformer-based architecture for diamond clarity grading.

F. Training Strategy

RAdam (Rectified Adam) is an enhanced version of the Adam optimizer developed to deliver more reliable and stable training, particularly in the initial step of optimization [26]. Although Adam adjusts learning rates utilizing second and first moment evaluations of gradients, it may suffer from aggressive or unstable updates at the start of training. RAdam overcame this problem by introducing a rectification system that automatically alters the adaptive learning rate depending on the difference of the gradients. The key concept in RAdam is to manage the adaptive stage dimensions utilizing a rectification term that is based on the training stage and the exponential decay rate of the second moment. The rectification factor is represented by:

$$\rho_t = \rho_\infty - 2 \cdot \frac{t \cdot \beta_2^t}{1 - \beta_2^t} \tag{18}$$

Here ρ_∞indicates the theoretical maximum value of ρ_t, $\rho_t$denotes the rectification term at step t, t refers to the current iteration step, and $\beta_2$represents the decay rate for the second moment estimate:

$$\eta_t = \begin{cases} \dfrac{\alpha}{\sqrt{\hat{v}_t} + \varepsilon} . \sqrt{\dfrac{\rho_t - 4}{\rho_\infty - 4}}, & if \rho_t \geq 5 \\ \dfrac{\alpha}{\sqrt{\hat{v}_t} + \varepsilon}, & if \rho_t < 5 \end{cases} \tag{19}$$

Here α signifies the base learning rate, η_t refers to the learning rate at step t, ϵ indicates a small constant used for numerical stability, and $\hat{v}_t$ points out the bias-corrected second moment estimate of the gradients. The model is trained using a learning rate of 1e-4 with the RAdam optimizer, ensuring stable convergence. It uses a weighted CrossEntropyLoss to handle class imbalance by assigning higher importance to minority classes. The training is carried out with a batch size of 32 over multiple epochs.

$$L = -\sum_{i=1}^{N} w_{y_i} \log(p_{y_i}) \tag{20}$$

Here, L denotes the total loss over N samples, p_{y_i} is the predicted probability of the true class y_i, and w i represents the class-specific weight by penalizing minority classes more strongly.

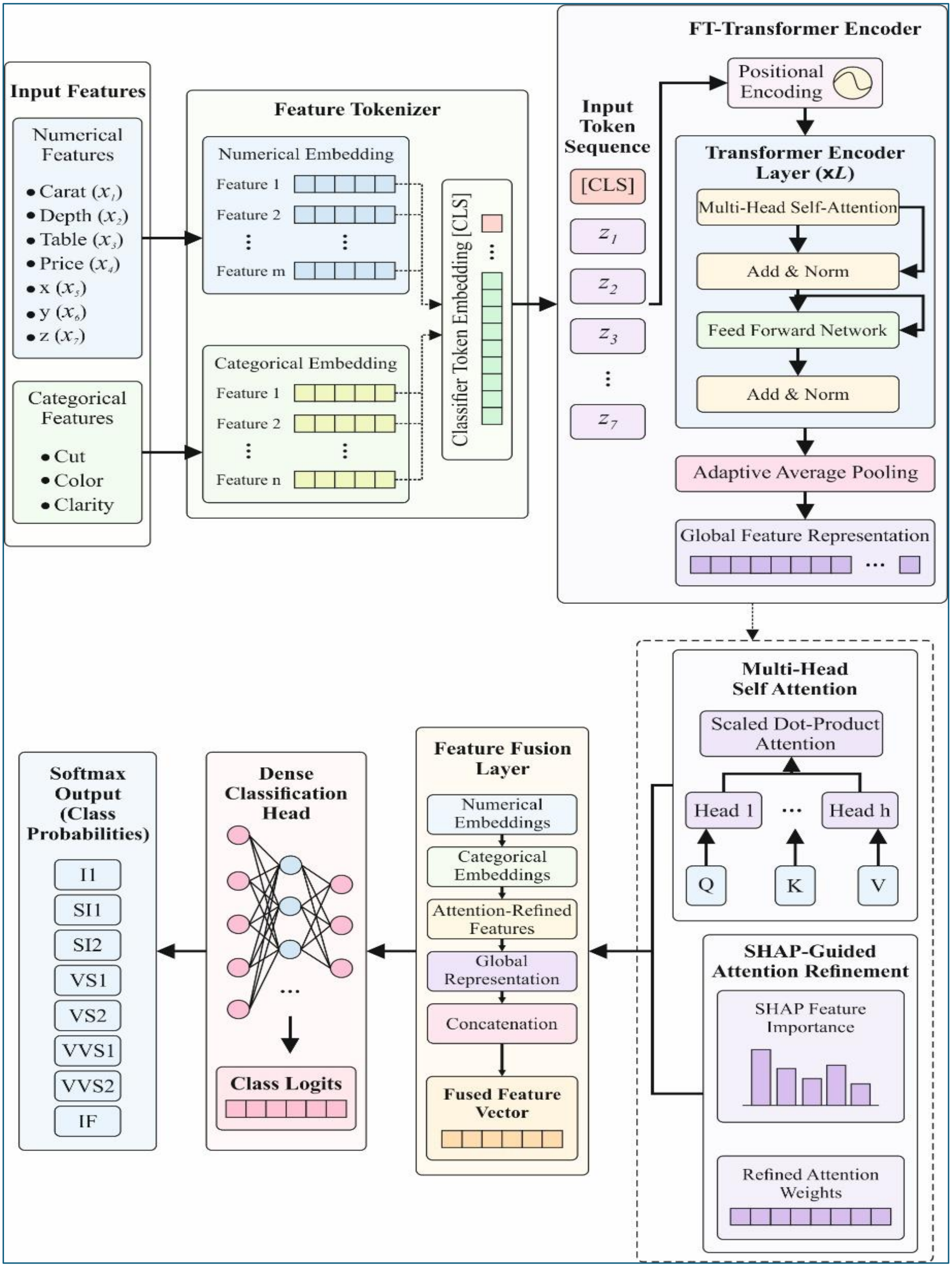

Fig. 4. Transformer-based architecture for diamond clarity grading.

IV. Experimental Analysis and Discussion

This section discusses the performance of the proposed SGT-Net model evaluated on the Diamond dataset. Fig. 5 displays the classifier outcome of the proposed SGT-Net model. The confusion matrix is presented in Fig. 5(a) using an 80:20 train-test split; the diagonal of the confusion matrix highlights strong predictions across all eight classes. The PR curve in Fig. 5(b) reflects a good balance between precision and recall, which demonstrates stable performance across all clarity classes. The ROC curve in Fig. 5(c) with high AUC values shows excellent classification capability. Therefore, the result indicates that the proposed SGT-Net model achieved strong prediction capability in diamond clarity grading.

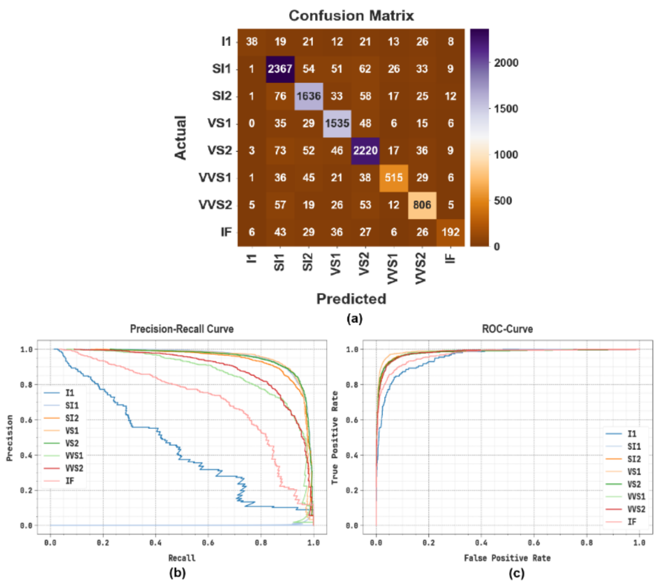

Fig. 5. Confusion matrix (a), PR (b), and ROC (c) Curve analysis of the SGT-Net model using an 80:20 train-test split.

Table II and Fig. 6 exhibit the performance evaluation of the proposed SGT-Net model using an 80:20 training and testing phase. The model achieves an average accuracy of 96.59% in the training phase and 96.57% in the testing phase, which indicates the model's strong generalization capability. The model also achieves excellent results in precision, recall, and F-score, which highlights the balanced classification performance of the model. Furthermore, the MCC and G-Measure values demonstrate strong predictive reliability and stable performance. Therefore, these findings validate that the proposed SGT-Net model is well-suited for accurate diamond clarity grading.

TABLE II. Performance Evaluation of SGT-Net Model Using 80:20 Training and Testing Phase

Metrics	Training Phase (80%)	Testing Phase (20%)
Accuracy	96.59	96.57
Precision	82.93	82.65
Recall	73.84	74.28
F-Score	76.52	76.75
MCC	75.40	75.55
G-Measure	77.40	77.56

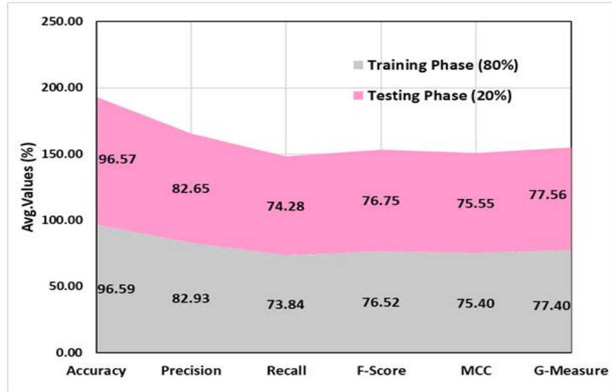

Fig. 6. Performance evaluation graph of the SGT-Net model using an 80:20 training and testing phase.

Fig. 7 presents the classifier result of the proposed SFT-Net framework. The confusion matrix is shown in Fig. 7(a) using a 70:30 train-test split; the diagonal of the confusion matrix shows that most samples are predicted correctly with very few misclassifications, which highlights the strong classification accuracy across all classes. The PR curve in Fig. 7(b) exhibits strong precision and recall across the majority of classes. The ROC curve in Fig. 7(c) further validates excellent discrimination between clarity grades. Therefore, the outcome confirms that the proposed SGT-Net model achieved robust and consistent performance in diamond clarity grading.

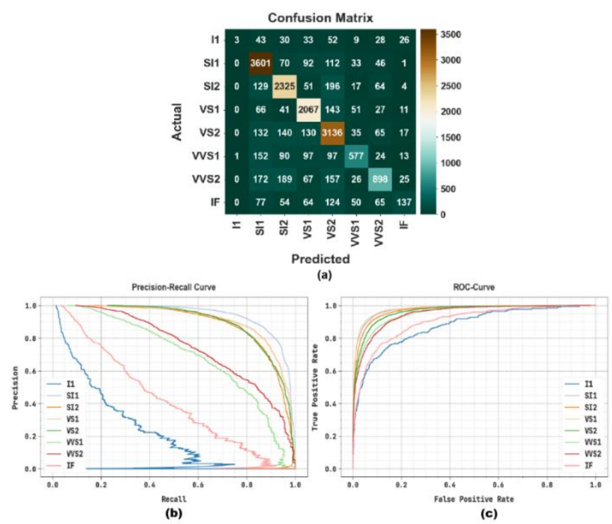

Fig. 7. Confusion matrix (a), PR (b), and ROC (c) Curve analysis of the SGT-Net model using a 70:30 train-test split.

TABLE III. Performance Evaluation of SGT-Net Model Using 70:30 Training and Testing Phase

Metrics	Training Phase (70%)	Testing Phase (30%)
Accuracy	94.76	94.69
Precision	71.63	74.83
Recall	60.66	60.62
F-Score	61.92	62.05
MCC	59.96	60.55
G-Measure	62.94	63.57

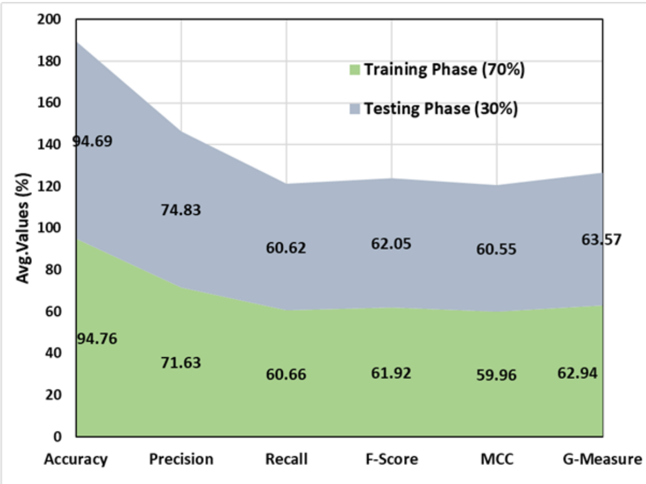

Fig. 8. Performance evaluation graph of the SGT-Net model using a 70:30 training and testing phase.

Table III and Fig. 8 show the performance evaluation of the proposed SGT-Net model using a 70:30 training and testing phase. The model attains an average accuracy of 94.76% in the training phase and 94.69% in the testing phase, which reveals the excellent generalization ability of the model. Additionally, the model attains well-balanced classification performance in metrics such as precision, recall, and F-score, which underlines the reliable classification performance of the model. The MCC and G-Measure values confirm the model's stability and robustness. Therefore, the obtained results highlight the effectiveness of the proposed SGT-Net model in accurately grading diamond clarity.

TABLE IV. COMPARATIVE ANALYSIS OF THE PROPOSED SGT-NET MODEL WITH EXISTING MODELS

Models	Accuracy	Precision	Recall	F1-Score	Inference Time(ms)
XGBoost	88.05	81.78	65.94	75.87	11.95
Stacking	83.71	80.44	72.90	67.00	9.620
LSTM	84.01	71.95	65.90	71.44	11.41
Transformer	85.42	74.32	67.69	68.45	10.35
RT- DETR- R18	79.90	76.91	66.86	69.31	16.80
FT-Transformer	92.84	79.36	69.52	72.81	9.42
TabTransformer	91.76	77.85	67.43	70.96	10.18
SAINT	93.21	80.14	70.68	74.02	11.37
NODE	90.93	75.62	65.87	68.54	8.76
TabNet	92.15	78.43	68.29	71.35	8.21
CatBoost	94.38	80.87	71.46	74.63	7.14
LightGBM	93.72	79.58	70.31	73.42	6.98
SGT-Net	**96.59**	**82.93**	**73.84**	**76.52**	**6.73**

Table IV signifies the comparative analysis of the proposed SGT-Net approach with existing methodologies [27, 28]. The outcome demonstrated that the RT- DETR- R18, Stacking, LSTM, and Transformer models had ineffective outcomes with accuracies of 79.90%, 83.71%, 84.015%, and 85.42%, respectively. Moreover, the XGBoost, NODE, TabTransformer, TabNet, FT-Transformer, and SAINT approaches achieved somewhat greater performance. In addition, the LightGBM and CatBoost methodologies achieved considerable results. At the same time, the proposed SGT-Net approach achieves promising performance with an accuracy of 96.59%, precision of 82.93%, recall of 73.84%, and F1-score of 76.52%, respectively. Thus, the proposed SGT-Net achieves superior classification performance compared with the existing methods, demonstrating its effectiveness for accurate diamond clarity grading.

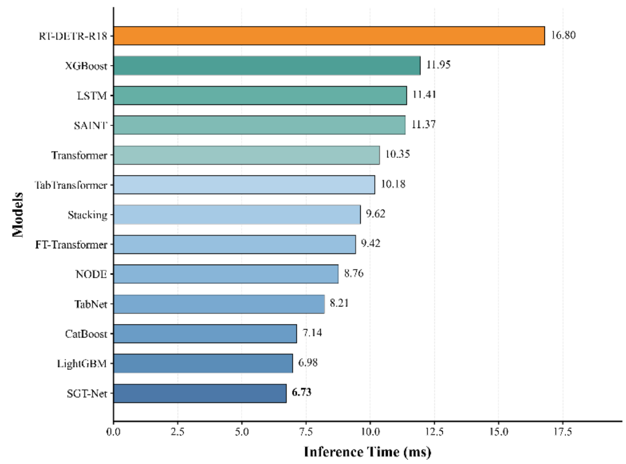

Fig. 9. Inference time of the SGT-Net model.

Fig. 9 signifies the inference time of the proposed SGT-Net model. Among the evaluated models, the XGBoost, Stacking, LSTM, Transformer, RT- DETR- R18, FT-Transformer, TabTransformer, SAINT, NODE, TabNet, CatBoost, and LightGBM approaches take longer inference times, whereas the proposed SGT-Net approach attains a lower inference time of 6.73 ms. Demonstrating its suitability for efficient and timely diamond clarity classification.

TABLE V. ABLATION STUDY OF THE PROPOSED SGT-NET MODEL

Configuration	Accuracy	Precision	Recall	F1-Score
Baseline	89.76	72.84	61.53	66.27
Transformer + MHSA	92.18	76.24	65.73	69.82
Transformer + MHSA + Feature Engineering	94.07	78.31	68.42	72.11
SGT-Net (SHAP-Guided Transformer Network)	**96.59**	**82.93**	**73.84**	**76.52**

Table V presents the ablation study of the proposed SGT-Net model. The Baseline configuration records the lowest performance, with an accuracy of 89.76%. Adding MHSA to the Transformer improves the results to 92.18% accuracy, while the inclusion of feature engineering further increases the accuracy to 94.07%. The proposed SGT-Net, which additionally incorporates SHAP-guided feature importance into the Transformer attention mechanism, achieves the highest performance with an accuracy of 96.59%, precision of 82.93%, recall of 73.84%, and F1-score of 76.52%, respectively. These results indicate that each added component contributes to the overall classification performance, with the SHAP-guided attention mechanism providing the final improvement in the proposed SGT-Net model.

TABLE VI. K-FOLD CROSS VALIDATION

Fold	Accuracy	Precision	Recall	F1 Score
Fold 1	96.21	82.14	72.31	75.18
Fold 2	96.84	83.47	74.26	76.93
Fold 3	96.43	82.61	73.18	75.89
Fold 4	97.02	84.16	75.02	77.81
Fold 5	96.58	82.94	73.71	76.41
Fold 6	95.97	81.72	71.84	74.73
Fold 7	96.76	83.28	74.48	77.12
Fold 8	96.35	82.38	72.96	75.67
Fold 9	97.18	84.52	75.63	78.24
Fold 10	96.61	83.05	73.89	76.58
Average	**96.60**	**83.03**	**73.73**	**76.46**
Std. Dev.	**±0.37**	**±0.85**	**±1.14**	**±1.10**

Table VI shows the 10-fold cross-validation analysis of the proposed SGT-Net model using accuracy, precision, recall, and F1-score. The model shows stable performance across all ten folds, with accuracy ranging from 95.97% to 97.18%. The highest accuracy is obtained in Fold 9, while the lowest is observed in Fold 6. The average accuracy, precision, recall, and F1-score across all folds are 96.60%, 83.03%, 73.73%, and 76.46%, respectively. The corresponding standard deviations are ±0.37, ±0.85, ±1.14, and ±1.10, indicating relatively small variations between the folds. These outcomes demonstrate the consistent performance and stability of the SGT-Net model across different data partitions.

Fig. 10 presents the SHAP-based explainability results of the proposed SGT-Net framework for diamond clarity grading. The figure provides both global and local explanations. The global explanation shows the overall importance of the features used for clarity grading, whereas the local explanation shows how individual features influence a particular prediction. This analysis helps to understand the model's decisions and provides greater clarity about the factors affecting the grading results.

The proposed SGT-Net combines feature engineering, SHAP-based feature prioritization, and Transformer learning for diamond clarity classification. SHAP helps identify the features that have a greater influence on the classification results, while the Transformer encoder with MHSA captures relationships among the different diamond attributes. The experimental results show that the model can effectively use these feature relationships to produce reliable grading results. Overall, SGT-Net provides a useful approach for automated diamond clarity assessment and related valuation applications.

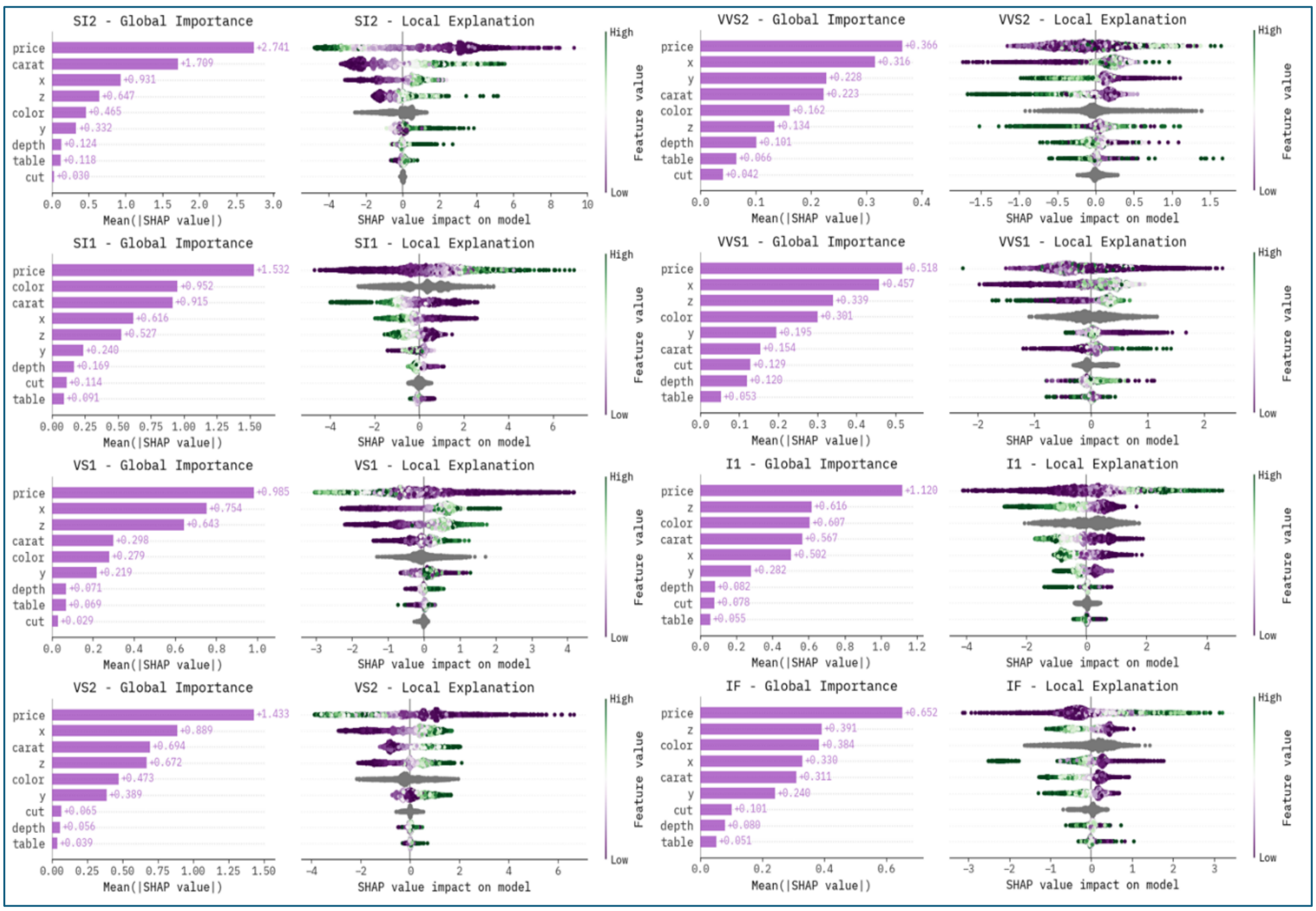

Fig. 10. SHAP-based analysis of the SGT-Net diamond clarity grading framework.

V. Conclusion and Future Works

This manuscript introduced an SGT-Net for diamond clarity grading. The proposed model includes a preprocessing stage that prepares the diamond grading dataset for effective and reliable model training. Besides, the feature engineering process is performed to enhance the representation of the dataset by deriving meaningful attributes from the raw diamond features. Moreover, SHAP-guided feature prioritization is utilized to recognize the most influential features in the data. The transformer-based architecture processes the data utilizing separate embeddings for numerical and categorical features, which are then passed through a transformer encoder with MHSA. In this setup, SHAP is used to understand feature importance and guide the model to pay more attention to the most relevant features during learning. The resulting fused feature representations are then passed through a dense layer followed by a softmax classifier to predict diamond clarity into 8 distinct classes. For training, the model is optimized utilizing the RAdam optimizer along with a weighted cross-entropy loss function. Experimental evaluation is carried out using performance metrics, demonstrating that the proposed SGT-Net effectively learns difficult feature relationships and enhances diamond clarity grading performance, making it suitable for resource valuation systems. Even though the proposed SGT-Net demonstrates better outcomes in diamond clarity grading, its precision relies on the quality and dissimilarity of the training data. Additionally, the transformer framework may need high computational resources and longer training time for huge datasets. In the future will concentrate on testing the system on large and more distinct diamond datasets from multiple sources. The integration of image characteristics and tabular data can be investigated to enhance grading accuracy. Furthermore, lighter transformer strategies and innovative explainability approaches may be examined to increase system efficacy and interpretability for real-time gemstone assessment and resource estimation applications.

Acknowledgment

The authors extend their appreciation to Northern Border University, Saudi Arabia, for supporting this work through project number (NBU-CRP-2026-1564). The authors are thankful to the Deanship of Graduate Studies and Scientific Research at the University of Bisha for supporting this work through the Fast-Track Research Support Program.

Data Availability Statement

The data that support the findings of this study are openly available in the Kaggle repository at: https://www.kaggle.com/datasets/joebeachcapital/diamonds/data [20].

References

[1] Huang, K., 2025, September. Comparative Analysis of Machine Learning and Deep Learning Models for Diamond Cut Grade Prediction. In Proceedings of the 2025 8th International Conference on Computer Information Science and Artificial Intelligence (pp. 1718-1723).

[2] Zhou, M., 2025, July. Enhancing Diamond Price Prediction through Machine Learning and Deep Learning: A Comparative Analysis of AGS and GIA Grading Systems. In Proceedings of the 2025 2nd International Conference on Big Data and Digital Management (pp. 439-445).

[3] Chang, X., Gao, W., An, J., Chen, H., Liu, Z., Qi, Z., Li, C., Ding, J., Li, W., Wang, K. and Wang, Y., 2025. Optical-grade diamond: characteristics, synthesis, and recent research progress. Functional Diamond, 5(1), p.2476690.

[4] Belfiore, E., Allinei, P.G., Houedry, P., Bahhi, M., Bartolacci, S., Saleh, A., Mosbah, M.B., Antoni, R., Lyoussi, A. and Groetz, J.E., 2025. Comparative assessment of feature extraction for fast neutron spectra prediction with machine learning algorithms using a CVD diamond detector. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, p.170951.

[5] Feng, Z., Dong, C., Xu, X., Liu, Y. and Wang, S., 2024. Deep learning-based automated optical inspection system for the additive manufacturing of diamond tools. 3D Printing and Additive Manufacturing, 11(6), pp.2045-2060.

[6] Kigo, S.N., Omondi, E.O. and Omolo, B.O., 2023. Assessing predictive performance of supervised machine learning algorithms for a diamond pricing model. Scientific Reports, 13(1), p.17315.

[7] Basha, M.S.A., Oveis, P.M., Prabavathi, C., Lakshmi, M.B. and Sucharitha, M.M., 2023, January. An efficient machine learning approach: Analysis of supervised machine learning methods to forecast the diamond price. In 2023 International Conference for Advancement in Technology (ICONAT) (pp. 1-6). IEEE.

[8] Liu, W., Li, J. and Gao, D., 2024. Failure Forensics of Shaped PDC Cutters Using Image Analysis and Deep Learning. SPE Journal, 29(04), pp.1832-1846.]

[9] Zhang, Q., Hardman, M.F., Stachel, T., Chinn, I., Seller, M., Kjarsgaard, B. and Pearson, D.G., 2025. A machine learning approach to single garnet geothermometry and application to tracing the fingerprint of superdeep diamonds. Geochemistry, Geophysics, Geosystems, 26(3), p. e2024GC012124.

[10] Abbas, M., Nazir, A. and Ali, U., 2025. Design, additive manufacturing, and machine learning prediction of multi-material diamond TPMS structure for improved mechanical performance. Progress in Additive Manufacturing, 10(11), pp.9793-9809.

[11] Wu, D., Xu, Z. and Guo, D., 2025. Machine learning accelerates programmable mechanics in isotropic diamond plate lattices. International Journal of Mechanical Sciences, p.110595.

[12] Yan, X., Yang, S., Zhang, S., Li, X. and Li, A., 2026. Diamond-DETR: lightweight real-time quality evaluation algorithm for synthetic diamonds. Scientific Reports.

[13] Zhang, S., Li, A., Ren, J. and Li, X., 2024. An enhanced YOLOv8n object detector for synthetic diamond quality evaluation. Scientific Reports, 14(1), p.28035.

[14] Mahmood, T. and Zia, A.W., 2025. Predicting the hardness of diamond-like carbon coatings using machine learning and generative adversarial networks. Journal of Manufacturing Processes, 149, pp.129-143.

[15] Triet, N.M., Khanh, V.H., Ngan, N.T.K., Khoa, T.D., Anh, N.T., Loc, V.C.P., Khiem, H.G. and Nam, T.B., 2024, November. Enhancing Diamond Grading and Certification Through Blockchain Technology: A Cross-Chain Analysis of EVM-Compatible Platforms. In International Conference on Metaverse (pp. 68-80). Cham: Springer Nature Switzerland.

[16] Mekala, R.R., Garratt, E., Muehle, M., Srinivasan, A., Porter, A. and Lindvall, M., 2024. Ai-guided defect detection techniques to model single crystal diamond growth. arXiv preprint arXiv:2404.07306.

[17] Yang, X., Feng, X., Cheng, C., Yu, J., Zhang, Q., Gao, Z., Liu, Y. and Chen, B., 2025. Automatic detection and classification of drill bit damage using deep learning and computer vision algorithms. Natural Gas Industry B, 12(2), pp.195-206.

[18] Basha, M.S.A., Sucharitha, M.M., Devi, A.U., Ashok, M. and Oveis, P.M., 2023, March. Exploring Machine Learning Models to Predict the Diamond Price: A Data Mining Utility Using Weka. In 2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1870-1877). IEEE.

[19] Randhawa, A.K., Sawarkar, A.D., Mundhe, S.B., Deshmukh, A.W. and Sonarkar, P.R., 2024, November. Predicting Diamond Prices Using Regression in Machine Learning. In International Conference on Paradigm Shifts in Communication, Embedded Systems, Machine

Learning and Signal Processing (pp. 164-176). Cham: Springer Nature Switzerland.

[20] https://www.kaggle.com/datasets/joebeachcapital/diamonds/data

[21] Alam, S., Ayub, M.S., Arora, S. and Khan, M.A., 2023. An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity. *Decision Analytics Journal*, *9*, p.100341.

[22] Devanathan, B., Varshitha, K.J., Kumar, L.P., Lakshmanan, S.A. and Prakash, N.K., 2025. Explainable AI Framework Using XGBoost with SHAP and LIME for Multi-Scale Household Energy Forecasting. *IEEE Access.*

[23] Wang, M., Yang, N. and Weng, N., 2023. Securing a smart home with a transformer-based IoT intrusion detection system. *Electronics*, *12*(9), p.2100.

[24] Wang, X., Zhu, Y., Wang, P., Luo, Y., Wei, M. and Wu, Y., 2025. Research on FT-transformer-based intelligent discrimination model for roof stability classification of retreating coal roadways. *Results in Engineering*, *28*, p.107115.

[25] Liao, W., Gao, C., Fang, J., Zhao, B. and Zhang, Z., 2024. A TCN-BiGRU density logging curve reconstruction method based on multi-head self-attention mechanism. *Processes*, *12*(8), p.1589.

[26] Dhingra, S., Gruosso, G. and Storti Gajani, G., 2025. A benchmark study of optimizers for short-term solar PV power forecasting using neural networks under real-world constraints. *Neural Computing and Applications*, *37*(29), pp.23909-23939.

[27] Huang, K., 2025, September. Comparative Analysis of Machine Learning and Deep Learning Models for Diamond Cut Grade Prediction. In *Proceedings of the 2025 8th International Conference on Computer Information Science and Artificial Intelligence* (pp. 1718-1723).

[28] Huo, X. (2026) 'Multimodal Diamond Inclusion Detection and Clarity Grading via Enhanced YOLOv7 with ECA and ASFF Modules,' *Informatica*, 50(8). https://doi.org/10.31449/inf.v50i8.12773.