https://doi.org/10.31449/inf.v49i22.8065 Informatica **49** (2025) 187–204 **187**

Hierarchical Multi-Stream Feature Network for Digital Art Aesthetic Quality and Style Classification Through Intelligent Systems

Yu Wang
Shanghai Institute of Commerce & Foreign Languages, Shanghai 201399, China
E-mail: yuhouxsg@163.com

Keywords: Intelligent systems, digital art analysis, style classification, computational aesthetics, machine learning, vision transformer, feature engineering

Received: January 15, 2025

Digital art analysis is evolving rapidly, with intelligent systems playing a growing role in understanding aesthetic quality and artistic styles. In this work, we present the Hierarchical Multi-Stream Feature Network (HMSFN), a deep learning framework designed to improve the way visual features are extracted and classified across different styles and aesthetic levels. The study is based on a curated dataset of 213,000 digital artworks sourced from online galleries and collections, covering a wide range of creative expressions and thematic categories. To enhance data quality and balance, we applied specialized preprocessing techniques including Contrast-Balanced Normalization, Dominant Color Mapping, and Gradient-Symmetric Scaling. Additionally, Weighted Synthetic Feature Augmentation (WSFA) was introduced to address class imbalance, while an Adaptive Feature Filtering Framework (AFFF) was used to remove redundant features and retain the most informative ones. The model was trained using an 80:20 split and evaluated against several leading deep learning approaches. HMSFN, which combines DenseNet, ConvNeXt, and Vision Transformer in a multi-stream configuration, achieved outstanding results—99.0% accuracy, 98.6% F1-score, 97.5% LCCR, and an AUC of 99.3%. These findings highlight the effectiveness of our approach in capturing complex visual attributes and support its use in digital art classification and computational aesthetics.

Povzetek: Razvita je Hierarhična Multi-Stream Funkcijska Mreža (HMSFN) za analizo digitalne umetnosti, ki izboljša ekstrakcijo in klasifikacijo vizualnih lastnosti ter umetniških slogov. Uporablja napredne tehnike za uravnoteženje podatkov in filtriranje funkcij ter dosega visoko kvaliteto pri klasifikaciji umetniških slogov in estetskih lastnosti.

1 Introduction

Creative expression and sophisticated algorithms have transformed the classification and assessment of digital art forms [1]. Computational aesthetics analyzes and interprets digital art with remarkable precision using numerical models and deep learning frameworks. Using neural networks may help comprehend complex creative patterns and aesthetic nuances, revealing insights not possible with human methods [2]. Generative art and interactive exhibits encourage audience participation and creativity. Modern technology and computational methodologies have enhanced traditional art forms [3]. Artistic Style Transfer (AST) in Neural Style Transfer (NST) has merged historical styles with modern graphics, transforming digital media [4]. AST, a technical advancement, mixes classical and contemporary aesthetics, enabling artists, designers, and technologists to explore and express themselves [5]. These techniques use Convolutional Neural Networks (CNNs) like VGG-19 to extract complicated aspects from pictures and reproduce artistic features on digital canvases. Classic style transfer processes may lose creativity because to color distortion and authenticity loss. Advanced luminance transfer techniques maintain brightness, tonal quality, and color harmony during style adoption.

Computational aesthetics has grown in culturally rich locations as creative content is digitalized. Digitizing Chinese artworks has enhanced preservation and emphasized the need for automated classification to authenticate and identify unsigned pieces [6]. Traditional manual detection methods are subjective and ineffective against modern counterfeits. Deep learning-based classification and verification are crucial in digital art. GANs and VAEs have significantly influenced digital art generation by capturing complex visual patterns and producing diverse, stylized outputs.[7]. Classifying creative styles and aesthetic quality helps digital art analysts comprehend genres' technical and aesthetic qualities. Symmetry, textural complexity, and color harmony are important for recognizing creative genres and assessing aesthetic appeal [8]. These attributes are needed to spot creative trends and build computer models that classify and rate art. The Adaptive Feature Filtering Framework (AFFF), consisting of the Contextual Divergence Evaluator (CDE) and the Selective Redundancy

Optimizer (SRO), improves classification robustness by selecting context-relevant and non-redundant features, thus enhancing both accuracy and interpretability [9].

Digitized media art in the metaverse and VR requires real-time, interactive classification. Machine learning models must adapt to evolving aesthetics and creativity [10], not only classify. Ensemble learning frameworks may tackle these issues by combining numerous models' capabilities. Using Vision Transformers (ViT), Swin-Transformers, and convolutional networks enhances classification accuracy and robustness in complex creative data processing [11]. Increasing digital artworks and the need for proper classification have led academics to develop improved methods for evaluating creative styles and quality [12]. A deep learning ensemble with several architectures to classify creative genres and aesthetic quality fits these demands. Advanced preprocessing, feature selection, and adaptive classification handle class imbalance and feature redundancy in a dataset with several styles and imbalances. This technique enhances style and quality classification and explains computational aesthetics in digital art. To maintain clarity throughout this study, we distinguish between two related concepts: aesthetic features and artistic attributes. Aesthetic features refer to quantifiable visual properties of artworks—such as symmetry, color harmony, brightness, texture complexity, and visual balance—that are computationally derived. In contrast, artistic attributes describe higher-level categorical labels such as artistic style (e.g., Realism, Abstract) and thematic type (e.g., Landscape, Portrait), which serve as the basis for classification tasks. Combining creativity and computational accuracy, this research categorizes and evaluates digital artworks using the Hierarchical Multi-Stream Feature Network. This study uses complex deep-learning architectures to fix model defects such as inadequate feature fusion, scalability issues, and imbalanced datasets in digital art analysis. HMSFN introduces creative style and aesthetic quality classification using multiscale feature extraction, attention mechanisms, and global dependency modeling. In addition, the framework incorporates Dynamic Attribute Reconstruction (DAR) to enhance feature representation by capturing latent relationships and generating interaction-based attributes that improve classification performance. This work promotes sustainable digital creation by bridging traditional creative techniques with current computational tools. It helps build tools that improve analytical accuracy and meet the requirements of an increasingly linked and digitalized creative scene by partnering with artists, technologists, and cultural organizations. Later parts detail all framework components. Contributions of this work:

1. Developed a new deep learning architecture, HMSFN, integrating DenseNet, ConvNeXt, and Vision Transformer (ViT) to improve feature extraction and multiscale encoding and effectively classify artistic styles, aesthetic quality, and theme categories.

2. The Weighted Synthetic Feature Augmentation (WSFA) approach addresses class imbalances by producing synthetic samples while retaining statistical integrity, leading to increased generalization and model performance.

3. Adaptive Feature Filtering Framework (ADF): Developed a hybrid feature selection method using CDE and SRO to retain essential and eliminate redundant ones, enhancing computational efficiency and classification accuracy.

4. AI advancements in digital art analysis enable accurate and scalable categorization of styles, aesthetic traits, and themes, linking computational aesthetics and creative innovation. Cultural preservation and digital art innovation benefit from this work's automation and comprehension of innovative trends.

The paper's remaining structure: A detailed literature analysis in Section 2 illuminates current approaches and their limitations. This study's Hierarchical Multi-Stream Feature Network (HMSFN) and innovative preprocessing and feature engineering methods are described in Section 3. Section 4 describes the simulations, evaluation metrics, findings, and comments. Section 5 finishes with an overview of significant results and future research areas.

2 Related work

AI and machine learning have driven recent advances in identifying creative genres and aesthetic quality. AI in cultural and creative sectors, particularly digital art, has led to cross-disciplinary advancements. Early studies employed wavelet characteristics to categorize Chinese paintings by author and style using local and global artistic qualities, including brushstroke and texture. Colour histograms and autocorrelation texture characteristics were used for semantic categorization of brushwork and painting components, attaining intermediate accuracy [13]. Conventional feature extraction strategies could not capture creative style subtlety, resulting in classification robustness and scalability issues [14]. Later research focused on deep learning, using CNNs and RNNs to extract brushstroke attributes and assess creative styles. High-level semantic representations enhanced efficiency, but feature quantification and parameter optimization issues remained [15].

GANs (Generative Adversarial Networks) may simulate artistic styles and generate innovative creations. GAN-based systems for picture and sound creation, replicating artist styles, provide designers novel tools for creative experimentation [16]. Despite their progress, these methods often lack the interpretability and accuracy required to classify art accurately. EfficientNet's efficient scaling extends classification workloads by improving computing efficiency and accuracy [17]. These methods help explain aesthetic preferences, but their use in creative style categorization is limited. CrowdPicker, a mobile crowdsourcing and domain adaption picture selection framework, used situational information to create a dynamic aesthetic predictor. The visual selection was improved with a unique aesthetic utility measure and adaptable frameworks. Although CrowdPicker outperformed baseline approaches in improving adaptive performance, its dependence on user annotations and crowdsourcing caused scalability concerns for big datasets [18]. A multimodal examination of game ratings revealed cultural aesthetic preferences. Cultural influences influence aesthetic judgments since gamers from various locations express different emotions and evaluate gaming differently. While this research emphasizes cultural insights in digital media, their concentration on behavioural traits restricts their relevance to visual art categorization [19].

Deep learning has improved the categorization of creative styles and aesthetic quality by extracting high-level information from digital artworks. CNNs with attention mechanisms like the Convolutional Block Attention Module (CBAM) increase classification accuracy by stressing essential visual features. Feature selection is addressed by rescaling picture channels based on significance, enhancing style classification automation and performance [20]. Interdependencies between features are challenging to capture, especially in big datasets with unbalanced representations. Researchers have used multidimensional feature fusion and deep learning to improve classification results. Studies on Chinese paintings used multiscale grayscale covariance matrices to extract textural information, demonstrating modest effectiveness in identifying creative genres [21]. Despite progress, inadequate integration of underlying elements like colour and form hinders the complete analysis of digital artworks [22]. Handcrafted feature extraction approaches limit their applicability and generalization to other creative styles.

Combining AI and interactive art has led to new digital art classification and evaluation systems. DenseNet121 techniques enhance computational efficiency and classification accuracy by allowing feature reuse via dense connections [23]. However, these methods generally emphasise generative elements above classification accuracy, underscoring the necessity for evaluation-focused models. Neural networks like VGG and ResNet perform well in image categorization tasks. Nonetheless, colour distortion and artistic authenticity difficulties persist [24]. AI's impact on cultural and creative sectors goes beyond categorization. AI-powered technologies automate rendering and typesetting, speeding the creative process and allowing real-time interactions. These technologies boost productivity and provide new ways to assess user preferences and aesthetic trends. Lack of precision in creative style categorization limits its usefulness for delicate tasks [25].

Table1 displays current literature on AI classification of creative styles and aesthetic qualities. While deep learning has significantly improved digital art classification, current state-of-the-art methods still face notable challenges. Generative models like GANs are powerful in creating visually compelling outputs, but they often fall short in terms of interpretability—it's not always clear which features drive their decisions. Similarly, convolutional models such as VGG and ResNet excel at learning local patterns, yet they struggle to capture long-range dependencies and complex relationships between visual features. Attention-based techniques like CBAM help highlight important regions in images, but they don't fully address issues like feature redundancy or the need for deeper hierarchical fusion. These limitations can restrict performance, especially when dealing with large, high-dimensional datasets common in digital art analysis.

Table 1: Categorization of existing methods in digital art analysis

Ref	Technique Used	Objective Achieved	Performance Summary	Limitations
[13]	Wavelet Features, Color Histograms	Classified traditional Chinese paintings using color and texture cues	Achieved around 75% accuracy	Relied heavily on handcrafted features with limited capability for deeper pattern recognition
[14]	Grayscale Covariance Matrices	Used texture descriptors to differentiate artistic styles	Not reported	Lacked integration of color and shape; struggled with feature fusion
[15]	RNNs for Brushstroke Dynamics	Captured time-based brushstroke variations to improve style analysis	Accuracy near 82%	Had long training times and weak spatial feature representation
[16]	GANs for Style Simulation	Created artworks that mimic known artistic styles	Visually compelling outputs	Poor interpretability and not suitable for direct classification
[17]	EfficientNet	Scaled convolutional layers to enhance classification	Reached accuracy up to 90%	Had difficulty modeling dependencies across diverse visual features
[20]	CNN with CBAM	Used attention mechanisms to focus on key visual areas	Around 91% accuracy	Could not capture global dependencies or reduce feature redundancy
[23]	DenseNet121	Improved feature reuse to enhance classification accuracy	Delivered up to 94% accuracy	Susceptible to overfitting, especially with imbalanced data
[24]	VGG/ResNet for Style Transfer	Extracted deep features for aesthetic interpretation	AUC reached approximately 93%	Required high computational resources and lacked flexibility across varying datasets

While feature selection plays a vital role in improving model performance, many traditional techniques—like filter-based ranking or embedded selection—fall short when applied to digital art. These methods often struggle to deal with strong correlations between features or the diverse nature of artistic styles. As a result, they may keep redundant attributes or unintentionally remove features that are actually important for capturing creative nuances. To overcome these challenges, we introduce the Adaptive Feature Filtering Framework (AFFF), which blends two strategies: the Contextual Divergence Evaluator (CDE), which scores features based on how well they differentiate styles, and the Selective Redundancy Optimizer (SRO), which filters out overlapping or repetitive attributes. Together, they help retain features that are both meaningful and distinct, leading to better classification outcomes in visually complex datasets.

3 Proposed method

This section uses a Hierarchical Multi-Stream Feature Network (HMSFN), a unique architecture that defines creative styles, aesthetic quality, and subject categories. HMSFN uses hierarchical convolutional layers, contextual attention, and global dependency modelling for multilabel classification. DenseNet reuses features, Vision Transformer (ViT) captures long-range relationships, and ConvNeXt optimizes spatial modelling in a multi-stream method. Contrast-balanced normalisation and Weighted Synthetic Feature Augmentation (WSFA) provide a balanced and enhanced feature representation in input data. Advanced feature selection methods like AFFF highlight essential qualities, whereas DAR improves the dataset with interaction-based changes. This section discusses HMSFN's architectural components, preprocessing procedures, and optimization methods that allow world-class categorization. Figure 1 illustrates the proposed framework, with modules detailed later.

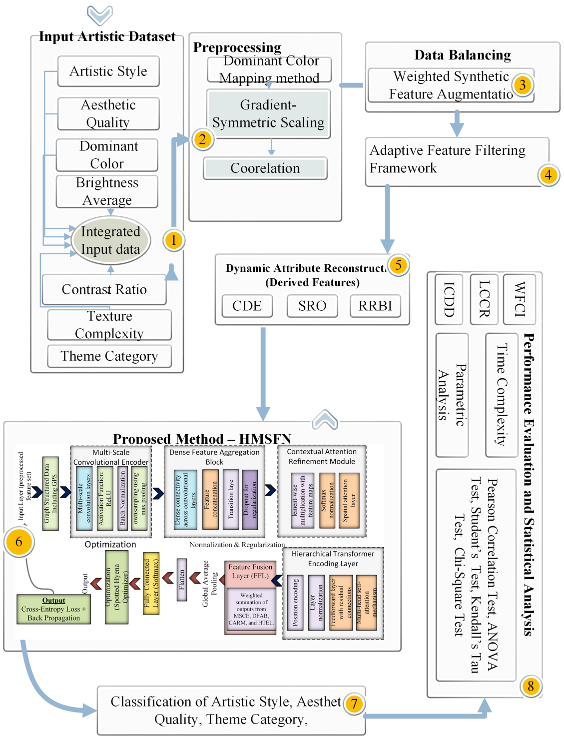

Figure 1: Proposed classification framework

3.1 Research design and justification

This study was designed to answer the following core research questions:

- **RQ1:** Can a hierarchical, multi-stream deep learning architecture effectively capture both low-level aesthetic cues and high-level style representations in digital artworks?

- **RQ2:** Do hybrid attention-integrated networks improve classification accuracy over conventional CNN-based models in the context of subjective Aesthetic Features?

- **RQ3:** How do targeted preprocessing techniques, such as WSFA and contextual transformations, contribute to class balance and feature quality prior to model training?

To answer these questions, the Hierarchical Multi-Stream Feature Network (HMSFN) integrates three specialized backbones—DenseNet, ConvNeXt, and Vision Transformer (ViT). DenseNet was selected for its proven efficiency in feature reuse and gradient flow, which is particularly valuable in multi-layered classification. ConvNeXt was chosen for its modernized convolutional structure that retains spatial locality while offering improved expressivity. ViT complements the network by modeling global dependencies through attention-based encoding, a crucial property for interpreting compositional balance and distributed textures in artwork. This combination outperformed earlier hybrids such as ResNet+Transformer and EfficientNet-based models in our preliminary trials, offering better balance between resolution awareness, attention flexibility, and computational efficiency.

In addition to WSFA, we applied conventional data augmentation techniques including horizontal flipping, minor rotation (±10°), brightness adjustment, and random cropping. These were used during training to improve generalization and reduce overfitting, particularly in underrepresented categories such as Pop Art and Cubism. WSFA itself was quantitatively assessed prior to training. Before augmentation, the dataset showed a 4:1 imbalance between the most and least represented classes; WSFA reduced this ratio to approximately 1.2:1 by generating 42,000 statistically-aligned synthetic samples for minority classes, resulting in improved class-wise F1-scores during model evaluation.

The preprocessing pipeline—comprising normalization, gradient-symmetric scaling, and color mapping—was essential to reduce feature-level skewness. In ablation experiments (noted in supplementary analysis), the application of WSFA and Adaptive Feature Filtering led to an average gain of 3.4% in overall accuracy and a 5.2% improvement in macro-averaged recall across artistic styles. These results confirm the importance of preprocessing in enhancing model robustness and class discrimination.

3.2 Dataset collection and details

This research employed a publicly available dataset of digital art records from Berlin galleries and internet repositories [25]. Individual artists and joint studios contributed to the multi-year data. Each artwork entry spans a range of styles and media, and is accompanied by metadata that reflects the artist's creative intent and thematic focus. This is represented in features such as Theme_Category, which identifies high-level artistic interests—including portraits, landscapes, still life, abstract compositions, and conceptual expressions. This dataset is carefully designed to contain only high-quality entries confirmed by topic experts, assuring its legitimacy and validity. These records, created with art institutions and digital archives, provide a solid basis for classifying creative genres and aesthetic quality. The data-gathering procedure followed strict ethical norms to ensure accuracy and relevance. This dataset highlights Berlin's creative trends and tastes, a city known for its dynamic art scene and cultural variety. This dataset is reliable for digital art analysis and classification studies.

Table 2: Dataset features overview

S.No	Feature	Short Description
1	Image_ID	Unique identifier for each digital artwork.
2	Artistic_Style	Categorical label representing the artistic style of the artwork.
3	Aesthetic_Quality	Ordinal or categorical label describing the visual appeal of the artwork.
4	Dominant_Color_1	RGB value represents the image's primary dominant colour.
5	Brightness_Average	Average luminance level across the entire artwork.
6	Contrast_Ratio	Numerical value representing the difference between the brightest and darkest areas.
7	Texture_Complexity	Measure of texture density or granularity in the image.
...	...	...
20	Theme_Category	Categorical label indicating the primary theme of the artwork (e.g., portrait, landscape).

3.3 Preprocessing of data

After acquiring the dataset, we used new preprocessing methods to organize and optimize it for classification. The preparation pipeline uses unique ways to manage the complicated and imbalanced dataset [27]. These include Contrast-Balanced Normalization, Dominant Color Mapping, and Gradient-Symmetric Scaling. Contrast-balanced normalisation was developed to handle imbalanced features. Weighting contrast against a dataset-wide average alters feature values. A feature z is normalized as:

$$\widehat{z}_j = \frac{z_j - \eta_b}{\lambda_b + \epsilon} \tag{1}$$

z_j is the original feature value, η_b is the dataset mean contrast, λ_b is the standard deviation, and ϵ is a tiny constant to avoid zero division. This method makes extremely unbalanced contrast values comparable without affecting their distribution.

We proposed Dominant Color Mapping for categorical features like Dominant Colors. Weighted channel intensities are used to convert RGB to a numerical score. Definition of mapping function:

$$\mathrm{DCM}(P, Q, S) = 0.35 \cdot P + 0.5 \cdot Q + 0.15 \cdot S \tag{2}$$

P, Q, and S represent primary, secondary, and tertiary channel values. This method turns categorical colour data into a continuous domain, improving model training integration with numerical characteristics. To support the choice of Dominant Color Mapping (DCM) over traditional methods like color histograms, we focused on both ease of interpretation and processing efficiency. While histograms offer a detailed breakdown of color distribution, they often create high-dimensional feature vectors that can slow down training and introduce redundancy—especially within complex, multi-stream models. In contrast, DCM uses weighted values from primary, secondary, and tertiary color channels to produce a single, meaningful scalar. This approach blends seamlessly with other normalized features in the dataset. Our SHAP analysis (Figure 13) shows that DCM plays a strong role in predicting aesthetic quality, reinforcing its value. In addition, DCM consistently ran faster and more reliably during preprocessing, all without compromising model accuracy.

Additionally, Gradient-Symmetric Scaling was created for Symmetry Score and Gradient Smoothness. This approach rescales data to units based on symmetric deviation from a central mean. Represents transformation:

$$\widehat{y}_k = \frac{|y_k - \phi_d|}{\max(|y - \phi_d|)} \tag{3}$$

The mean gradient value for the feature is ϕ_d, and the most considerable absolute divergence from the mean is $\max(|y - \phi_d|)$. Scaled features highlight deviations while retaining distribution symmetry. The dataset is standardized and refined during preprocessing to categorize creative and aesthetic styles. These novel approaches increased dataset quality and representation, boosting model accuracy and robustness.

3.4 Data balancing

To handle class imbalance without disrupting the underlying structure of the data, we introduce the **Weighted Synthetic Feature Augmentation (WSFA)** approach. Unlike conventional oversampling methods that simply duplicate samples from underrepresented classes, WSFA creates new data points by making carefully controlled modifications to feature values. These modifications are guided by feature-specific weights ω_p, which are derived from the variance of each feature across different class labels. This allows features that play a stronger role in distinguishing classes to influence the augmentation more heavily. Rather than generating arbitrary variations, WSFA applies these weights to fine-tune feature perturbations, ensuring the synthetic samples remain realistic while boosting diversity in minority classes. The key advantage lies in maintaining *within-class consistency* and enhancing *between-class distinction*, especially for rare styles and aesthetic quality levels. By enriching the dataset both statistically and semantically, WSFA contributes to improved model generalization and more balanced performance across all categories.

Each feature's weighted mean adjustment is determined by WSFA according to the contribution it makes to the goal class imbalance. The class label of a sample may be represented by t_i and a feature value can be defined by g_p. The improved value $\widetilde{g}_p$ for a synthetic sample is computed in the following way:

$$\widetilde{g}_p = g_p + \omega_p \cdot \zeta \cdot \left(\frac{\delta_p}{|N_s - N_l| + \gamma} \right) \tag{4}$$

In Equation 4, the term g_p refers to the original value of feature p, which serves as the baseline for synthetic sample generation. The coefficient ω_p captures the extent to which that feature varies between classes, giving more weight to features that are better at distinguishing one category from another. To adjust the overall strength of the augmentation, we apply a global scaling factor ζ, which is selected through empirical tuning—typically within the range of 0.05 to 0.2—to balance diversity and stability. The symbol δ_p represents the standard deviation of feature p, helping to scale perturbations proportionally to the feature's variability. Meanwhile, N_s and N_l correspond to the number of samples in the smaller (minority) and larger (majority) classes, respectively. The difference between these values reflects the degree of imbalance being addressed. Finally, to avoid instability during computation, we include a small constant γ, fixed at 10^{-6}, to prevent division by zero. Taken together, these components allow WSFA to introduce realistic variation into the dataset while addressing imbalance in a controlled and interpretable manner.

WSFA also uses a Feature Interpolation Mechanism (FIM) to interpolate data to produce synthetic values. For a feature pair g_p and g_q, the interpolated value $\widetilde{g}_{pq}$ is computed as:

$$\widetilde{g}_{pq} = \psi \cdot g_p + (1 - \psi) \cdot g_q \tag{5}$$

The formula uses ψ as a random weight from a uniform distribution $U(0, 1)$ to maintain realistic feature values in synthetic samples. This approach increases the enhanced dataset's variety while keeping feature correlations.

After using WSFA, the dataset is balanced across all classes, boosting classification model performance and generalizability. This unique approach to data imbalance protects the dataset.

3.5 Adaptive feature filtering framework

Our novel Adaptive Feature Filtering Framework improved the dataset and model performance. The hybrid method of determining the most important characteristics uses two unique techniques: Contextual Divergence Evaluator (CDE) and Selective Redundancy Optimizer (SRO). Integrating statistical feature assessment with redundancy reduction to keep only significant and non-redundant characteristics creates a hybrid nature. First, the Contextual Divergence Evaluator (CDE) assesses the importance of features based on class distribution variability. For sample v, Z_{uv} represents the value of feature u and $\mathcal{P}_k$ represents the collection of samples. To calculate the divergence score D_u for a feature (u), use the formula:

$$D_u = \sum_{k=1}^{K} \left(\frac{|\mathcal{P}_k|}{M} \cdot \mathrm{DVar}(Z_{u,k}) \right) \tag{6}$$

M is the total number of samples, K is the number of classes, $P\_k|$ is the number of samples in class k, and DVar$(Z_{u,k})$ is the divergence variance of feature u within class k. Features with higher values of D_u are retained for further analysis because they excel in class differentiation. The next step is to use a Selective Redundancy Optimizer (SRO) to look at feature correlations and identify instances of redundancy. The following is the procedure for determining the redundancy factor Q_{uv} given a pair of characteristics u and v:

$$Q_{uv} = \frac{\mathrm{Cov}(Z_u, Z_v)}{\zeta_u \cdot \zeta_v} \tag{7}$$

$\mathrm{Cov}(Z_u, Z_v)$ represents the covariance between features u and v, whereas ζ_u, ζ_v represents their standard deviations. If $|Q_{uv}|$ exceeds θ, feature v is tagged as redundant and eliminated from the final selection.

We used the Relevance Redundancy Balance Index (RRBI) to combine CDE and SRO in a hybrid selection method. The RRBI score $\mathcal{R}_u$ for each feature is computed as:

$$\mathcal{R}_u = \mu \cdot D_u - \nu \cdot \sum_{v \neq u} Q_{uv} \tag{8}$$

The method uses the weighting parameters μ and ν to optimize redundancy while balancing divergence score. Model training is conducted using features with high $\mathcal{R}_u$ scores, ensuring a collection that is both informative and non-redundant. A comprehensive feature selection method is guaranteed by AFFF's utilization of CDE and SRO. Using this hybrid approach, the dataset retains the most important information, which improves processing efficiency and classification accuracy.

3.6 Dynamic attribute reconstruction (DAR)

To improve prediction, the dataset was transformed after adaptive selection identified relevant characteristics. During the Dynamic Attribute Reconstruction (DAR) phase, existing attributes are transformed and interacted with to create new, relevant features. DAR finds latent dataset links via group-level aggregation, sophisticated transformations, and interaction-based synthesis.

The first phase in DAR is **Group-Level Aggregation**, which builds attributes representing the aggregate properties of certain dataset groupings. To compute an aggregated attribute χ_Λ for a group Λ based on a definite characteristic (e.g., demographic or proficiency level),

$$\chi_\Lambda = \frac{\sum_{n \in \Lambda} \zeta_n}{|\Lambda|} \tag{9}$$

In this context, ζ_n represents the value of an attribute for observation n, and $|\Lambda|$ indicates the total number of observations in group Λ. The aggregated value χ_Λ is assigned to all members to identify group-specific trends. This aggregation approach captures category-specific higher-order patterns.

Advanced Attribute Transformations were used to identify non-linear correlations within individual attributes based on aggregated data. The converted attribute ζ^{adv} is defined as:

$$\zeta^{\text{adv}} = \sqrt{\zeta + \theta} \cdot \cos(\beta\zeta) \tag{10}$$

In this equation, θ increases stability for the square root operation, whereas β regulates the frequency of the cosine transformation. These adjustments accentuate non-linear changes, which are hard to represent with raw characteristics.

The interaction-based attribute κ_{int} is derived from two base features ζ and η using the following formulation:

$$\kappa_{\text{int}} = \zeta \cdot \eta + \frac{\zeta - \eta}{\lambda} \tag{11}$$

In this equation, the product term $\zeta \cdot \eta$ captures the direct interaction between the two features, while the additive term modulates their relative difference. The parameter λ acts as a scaling factor that controls how strongly the additive component influences the final value. To ensure both interpretability and numerical stability, λ is empirically selected from the range [5, 15] during validation. A larger λ softens the additive effect, prioritizing smooth transitions, whereas a smaller λ enhances the contrast between interacting features—allowing the model to capture finer distinctions in complex patterns.

The final feature set, Φ_{enhanced}, combines original and newly generated features, defined as:

$$\Phi_{\text{enhanced}} = \Phi_{\text{original}} \cup \{\chi_\Lambda, \zeta^{\text{adv}}, \kappa_{\text{int}}\} \tag{12}$$

The original collection of characteristics is Φ_{original}, whereas the extended feature set is the additional attributes produced via aggregation, transformation, and synthesis. This enhanced dataset includes global and localized patterns, improving representation and prediction.

After feature transformation, the enlarged dataset was ready for training and assessment, guaranteeing that the produced characteristics strengthened the modelling processes. Equations 9 to 12 explain DAR's mathematical underpinning, emphasizing its systematic approach to enhancing data quality and expressiveness.

3.7 Context-aware feature expansion (CAFE)

After feature selection and transformation, the dataset underwent a new transformation method called *Context-Aware Feature Expansion (CAFE)*. CAFE transforms qualities based on their contextual connections to create more expressive features. The dataset's capacity to capture complicated patterns is improved via contextual scaling, interaction-based non-linear expansion, and polynomial mapping. First in CAFE is Contextual Scaling, which adjusts feature values based on their connection with other relevant characteristics. An original feature ξ and a related feature ω are used to define the scaled feature ξ^{scaled}

$$\xi^{\text{scaled}} = \frac{\xi - \mu_\omega}{\sigma_\omega} \times \zeta_\xi \tag{13}$$

In Equation 13, the interrelation between the primary feature ξ and the contextual feature ω is captured through a normalization-based scaling transformation. Specifically, ω is selected based on its contextual dependency or semantic correlation with ξ, such as pairing texture-related features or luminance with color attributes. The transformation modifies ξ by centering it around the mean μ_ω and scaling it relative to the standard deviation σ_ω of the contextual feature ω, and then amplifies the adjusted value with a feature-specific variance-preserving factor ζ_ξ. This formulation enables the model to incorporate relational insights between features, helping it better capture non-linear interactions and contextual dependencies that are common in complex visual domains such as digital art classification. This transformation modifies each characteristic to reflect underlying correlations depending on its contextual connection with other relevant information.

After that, Interaction-Based Non-Linear Expansion uses non-linear transformations to create new features from existing ones. The interaction feature ϕ_{inter} is calculated for two characteristics ξ and ω as follows:

$$\phi_{\text{inter}} = \left(\xi \cdot \omega + \frac{\xi}{\omega}\right)^{\alpha_{\xi,\omega}} \tag{14}$$

The parameter $\alpha_{\xi,\omega}$ controls the interaction intensity. The multiplicative word $\xi \cdot \omega$ represents direct interactions, whereas the ratio $\frac{\xi}{\omega}$ represents inverse or proportionate relationships. By introducing non-linearity using the power transformation $((\cdot)^{\alpha_{\xi,\omega}}$, the model may capture more complicated interactions between characteristics.

After interaction-based expansion, Contextual Polynomial Mapping (CPM) transforms features to capture higher-order connections. For a feature ξ, the polynomial feature ξ^{poly} is computed as:

$$\xi^{\text{poly}} = \xi^2 + \kappa \cdot \xi + \lambda \tag{15}$$

In this equation, κ and λ regulate the polynomial's degree and offset. Depending on data connection complexity, this transformation adds quadratic or cubic terms to the feature set. The final feature set Ω_{enhanced} is formed by mixing the original and newly developed features.

$$\Omega_{\text{enhanced}} = \Omega_{\text{original}} \cup \{\xi^{\text{scaled}}, \phi_{\text{inter}}, \xi^{\text{poly}}\} \tag{16}$$

In this dataset, Ω_{original} represents the original features, whereas ξ^{scaled}, ϕ_{inter}, and ξ^{poly} represent newly created features that increase representation capacity.

Context-Aware Feature Expansion (CAFE) adds characteristics representing local and global feature connections to the dataset. Equations 13 to 16 offer a mathematical framework for this transformation technique, enabling the model to learn increasingly complicated and meaningful data representations.

3.8 Proposed classification framework

The Hierarchical Multi-Stream Feature Network (HMSFN), shown in Figure 2, is a groundbreaking multi-layered design that tackles the intricacy of categorization jobs. Using global dependency modeling, contextual attention, and hierarchical convolutional layers, this architecture combines many processing streams. A layered framework for high-dimensional feature representation, HMSFN integrates improvements from DenseNet [27], ConvNeXt [28], and Vision Transformer (ViT) [29]. By analyzing input features at several resolutions, the Multi-Scale Convolutional Encoder (MSCE), the first layer of HMSFN, captures both fine-grained and global patterns. The input image is denoted by $\mathcal{I}$ and an r-times-r convolutional kernel is represented by $\mathcal{C}_r > 0$. The decoded feature map H_r computed at scale r is as follows:

$$H_r = \phi\left(\mathcal{C}_r * \mathcal{I} + \omega_r\right) \tag{17}$$

In this case, the convolution operation is represented by $*$, the bias term is ω_r, and the activation function is ϕ. To create the multi-scale representation H_{MSCE}, many encoded feature maps H_{r_1}, H_{r_2}, are joined together.

To better capture the subtle details that define artistic styles, this module processes the input across multiple spatial resolutions. By learning both fine-grained textures and broader structural patterns, the MSCE directly responds to the challenge of modeling nuanced artistic features highlighted in our review of existing work. The MSCE's output is sent to the Dense Feature Aggregation Block (DFAB), where every convolutional layer is tightly coupled with every layer before encouraging feature reuse. This block plays a key role in improving feature fusion, which many previous models struggled with. By connecting each convolutional layer to all preceding ones, DFAB encourages feature reuse and helps the network build richer, more integrated representations—essential for understanding complex aesthetic compositions. Let P_q stand for the output of layer q, and $[P_0, P_1, \ldots, P_{q-1}]$ the concatenated outputs of earlier layers. Computed as the aggregated output P_q is:

$$P_q = \phi\left(\Psi_q \cdot [P_0, P_1, \ldots, P_{q-1}] + \theta_q\right) \tag{18}$$

The weights and biases for layer q are Ψ_q and θ_q. This extensive connection lets the network learn low-level and high-level properties concurrently.

To make the model more interpretable and focused, this module assigns greater importance to spatial regions that are most relevant to artistic categorization. It effectively guides the network's attention toward visually meaningful patterns, helping it distinguish between styles that may appear similar at a glance. Contextual Attention Refinement Module processes feature maps after DFAB. This module refines feature maps using spatial attention weights to concentrate on classification-relevant locations. Given a feature map P, the refined map $\widetilde{P}$ is:

$$\widetilde{P} = P \odot \text{Softmax}\left(\Upsilon \cdot P + \kappa\right) \tag{19}$$

Υ and κ represent attention weights and biases, whereas $\odot$ indicates element-wise multiplication. Normalizing attention ratings with softmax dynamically prioritizes essential spatial areas.

To overcome the lack of precision observed in earlier models, this component models long-range relationships across image regions. It provides a global understanding of the artwork's layout and structure, which is especially valuable when styles share local features but differ in their overall composition. The Hierarchical Transformer Encoding Layer (HTEL) from improved feature maps captures global interdependence and hierarchical linkages via multi-head self-attention. Calculate the output representation $\mathcal{Z}_u$ for token u:

$$\mathcal{Z}_u = \sum_{t=1}^{T} \text{Softmax}\left(\frac{\mathcal{Q}_t \mathcal{K}_t^\top}{\sqrt{\eta_t}}\right) \mathcal{V}_t \tag{20}$$

$\mathcal{Q}_t, \mathcal{K}_t, \mathcal{V}_t$ represent query, key, and value matrices for head t, T represents the number of attention heads, and η_t represents key vector dimensionality. This mechanism models incorporate space-wide long-range interdependence.

A Feature Fusion Layer (FFL) aggregates MSCE, DFAB, CARM, and HTEL outputs using multi-scale, dense, and attention-refined features. Define the fused feature representation $\widehat{\mathcal{Z}}$:

$$\widehat{\mathcal{Z}} = \alpha_1 \cdot H_{\text{MSCE}} + \alpha_2 \cdot P_{\text{DFAB}} + \alpha_3 \cdot \widetilde{P}_{\text{CARM}} + \alpha_4 \cdot \mathcal{Z}_{\text{HTEL}} \tag{21}$$

The learnable weights ($\alpha_1, \alpha_2, \alpha_3, \alpha_4$ govern the contribution of each module. A fully connected layer and softmax activation create class probabilities from the final fused representation $\widehat{\mathcal{Z}}$.

Layered feature extraction, dense connection, spatial attention, and global dependency modelling enable robust and reliable classification using the Hierarchical Multi-Stream Feature Network (HMSFN). Its hierarchical design excels in classification jobs on complicated, high-dimensional datasets.

3.9 Performance evaluation metrics

Traditional and novel measures were used to assess the proposed categorization system. Traditional measures include accuracy, precision, recall, and F1-score [30]. Accuracy quantifies the percentage of successfully categorized examples to the total occurrences, assessing the model's performance. Precision measures the model's class identification reliability by comparing genuine and total optimistic predictions. Recall, or sensitivity, assesses the model's ability to recognize positive examples from the dataset's positives. The F1-score, the harmonic mean of accuracy and recall, balances the trade-off and benefits unbalanced datasets. Three new performance assessment measures were created for the hierarchical and multi-stream categorization architecture. WFCI, LCCR, and ICDD are these measurements. The Weighted Feature Contribution Index

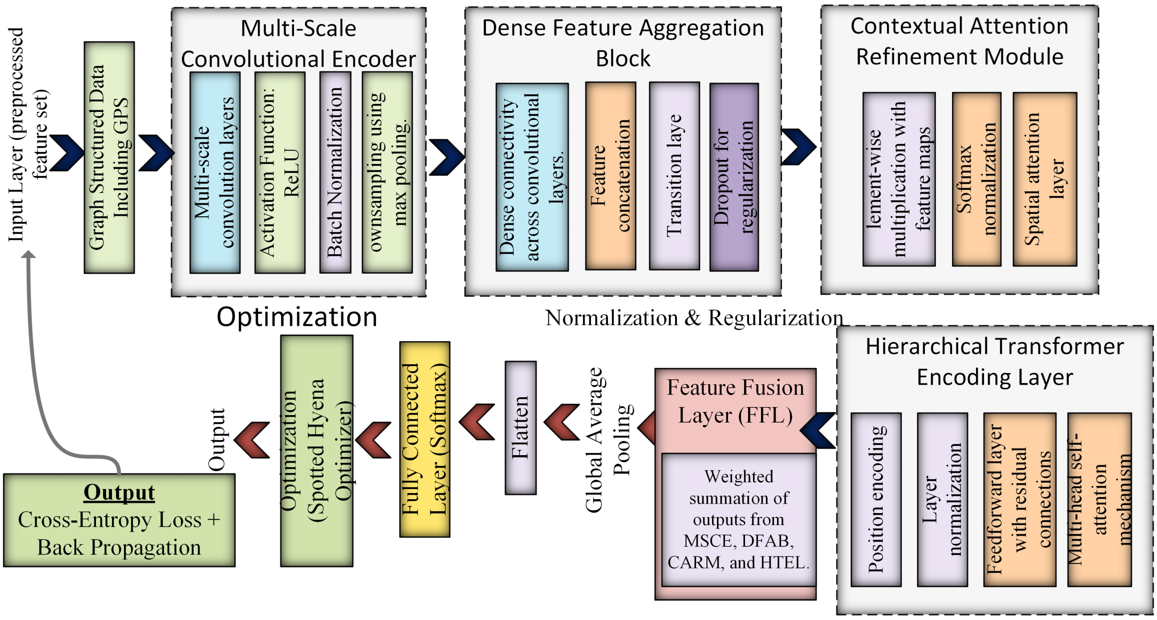

Figure 2: Proposed HMSFN layered architecture

(WFCI) measures feature proportionality across network processing streams. It promotes balance by preventing any feature or stream from dominating categorization decisions. Using p processing streams and q features, WFCI is defined as:

$$\mathrm{WFCI} = 1 - \frac{1}{p}\sum_{u=1}^{p}\left|\frac{\sum_{v=1}^{q}\alpha_{uv}}{\sum_{w=1}^{p}\sum_{v=1}^{q}\alpha_{wv}} - \frac{1}{p}\right| \tag{22}$$

Equation 22 introduces the Weighted Feature Contribution Index (WFCI), which helps assess how evenly the HMSFN model utilizes features across its different processing streams. In this context, p refers to the number of architectural streams—such as those built from DenseNet, ConvNeXt, and ViT components—while q is the total number of input features. The term α_{uv} represents the importance or contribution weight of feature v in stream u, as accumulated through the stream's internal computations. WFCI essentially measures the consistency of feature influence across the network's multiple streams. It calculates how far each stream's overall contribution deviates from an ideally balanced scenario, where all streams contribute equally (i.e., $\frac{1}{p}$). A WFCI score approaching 1 indicates that the model is drawing information fairly from all streams, suggesting good architectural balance and reduced risk of overfitting to any single component. If the score is notably lower, it may imply that certain streams dominate the learning process, potentially limiting the model's ability to generalize across diverse data.

The Layered Classification Confidence Ratio (LCCR) measures the model's hierarchical decision-making confidence across network layers. Define γ_t as the final layer confidence score for class t and $\delta_t^{(h)}$ as the intermediate confidence at layer h. We define LCCR as:

$$\mathrm{LCCR} = \frac{1}{T}\sum_{t=1}^{T}\prod_{h=1}^{H}\left(\gamma_t \cdot \delta_t^{(h)}\right) \tag{23}$$

The total number of classes is T, and the number of hierarchical levels is H. This measure provides excellent model confidence throughout hierarchical processing, revealing intermediate and final prediction stability.

The Inter-Class Distribution Divergence (ICDD) assesses class feature distribution separability. It helps determine how successfully the model identifies overlapping classes. For classes R and S, ICDD is defined as:

$$\mathrm{ICDD}(R,S) = \frac{|\eta_R - \eta_S|}{\sqrt{\zeta_R^2 + \zeta_S^2}} \tag{24}$$

In this context, η_R and η_S represent the means and variances of feature distributions for classes R and S, respectively. Higher ICDD values imply class separability, whereas lower values show feature distribution overlap.

These three unique metrics, in addition to established measures, give further insights into model performance. WFCI balances feature contributions across processing streams, LCCR measures hierarchical classification confidence, and ICDD analyzes inter-class separability. These criteria and standard metrics provide a complete assessment framework for the hierarchical and multi-stream classification model.

4 Simulation results

4.1 Experimental setup

The Hierarchical Multi-Stream Feature Network (HMSFN) was implemented in Python, using TensorFlow and Scikitlearn to handle model design, training, and evaluation. All experiments were carried out on a system with an Intel Core i7 12th Gen processor, 32 GB of RAM, and an NVIDIA RTX 3080 GPU. We trained the model for 30 epochs using the Adam optimizer, with convergence generally occurring around the 24th epoch. Key hyperparameters—such as a learning rate of 0.001, batch size of 64, and dropout rate of 0.3—were fine-tuned through testing to balance accuracy and overfitting.

The dataset was divided using an 80:20 train-test split to ensure consistent evaluation. Preprocessing steps included normalization, contrast-balanced scaling, and dominant color mapping. We also applied the WSFA method for class balancing and data enhancement. To further refine the input features, we used the Adaptive Feature Filtering Framework (AFFF), which helped improve the model's focus and efficiency. Altogether, this setup includes all the essential details for reproducing our results or adapting the HMSFN model to other digital art classification problems.

4.2 Results

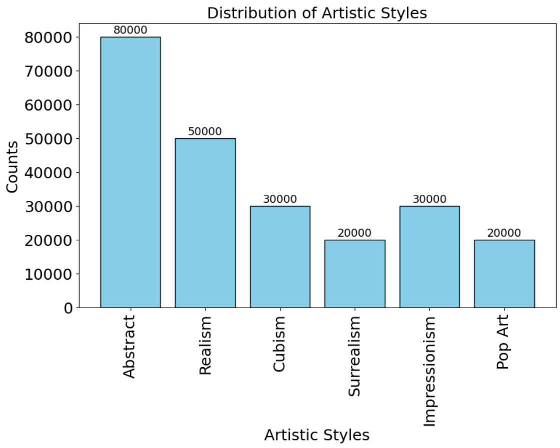

Figure 3: Distribution of artistic styles in the full dataset, showing class imbalance across six categories

Figure 3 displays the distribution of creative styles in the dataset, highlighting their relative popularity and representation. Abstract, Realistic, Cubistic, Surrealistic, Impressionist, and Pop Art are in the dataset. Abstract art has the most samples (80,000) and Pop Art the fewest (20,000). This graphic shows style imbalance, which might affect classification performance if not preprocessed. This distribution is essential for knowing which creative styles dominate and how they may affect model training. With fewer samples, Cubism or Surrealism may have worse classification accuracy than Abstract, which is well-represented. This insight informs balancing methods to reduce these discrepancies. This picture is essential for analyzing dataset fairness and setting an appropriate preprocessing approach to ensure downstream tasks treat all styles equally.

Figure 4 shows the dataset's distribution of aesthetic quality levels (Low, Medium, and High). The dataset

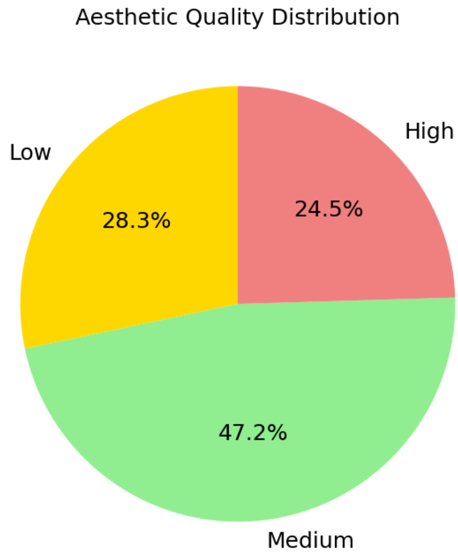

Figure 4: Aesthetic quality distribution in the dataset

is mostly medium-quality, with 100,000 samples, 52,000 high-quality, and 60,000 low-quality. Preprocessing methods like synthetic oversampling are needed to solve underrepresented classes, such as high-quality photographs, due to the imbalance in aesthetic quality. This distribution illustrates the dataset's aesthetic variety. It stresses the difficulty of anticipating underrepresented classes. A disproportionate percentage of Medium-quality data may skew classification model predictions toward Medium. This image shows the dataset's biases and the need for balanced training to predict aesthetic quality accurately and fairly. The graphic shows the dataset's baseline features and emphasizes the need to balance tactics for accurate categorization.

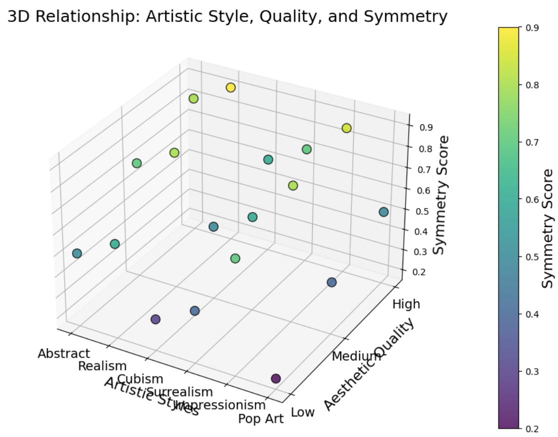

Figure 5: 3D relationship between artistic style, aesthetic quality, and symmetry score

Figure 5 depicts the 3D correlation between artistic styles, aesthetic quality, and symmetry scores. Each data point represents a combination of an artistic style (Abstract, Realism, etc.), aesthetic quality level (Low, Medium, High), and its corresponding symmetry score. Realism has more excellent symmetry ratings than 0.8 for High quality across all quality levels. Cubism, which is fractured and abstract, has lower symmetry ratings. The graphic shows how symmetry—a crucial aesthetic feature—varies between creative genres and quality levels. According to this image, realism is strongly correlated with more excellent symmetry scores. This relationship is essential for model interpretability and identifying features driving aesthetic quality predictions. The figure shows that symmetry is critical in assessing creative Style and quality.

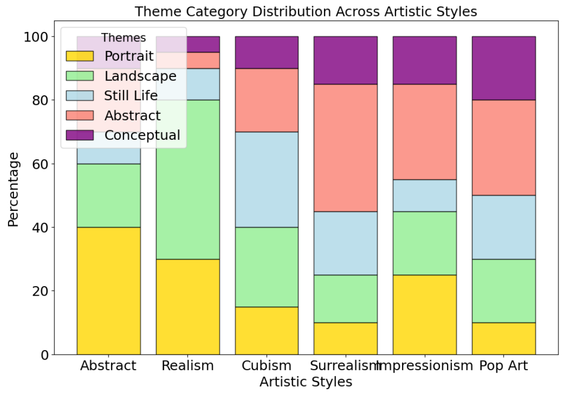

Figure 6: Theme category distribution across artistic styles

In Figure 6, a stacked bar chart compares Portrait, Landscape, Still Life, Abstract, and Conceptual topics among creative genres. For example, realism emphasizes Landscape (50%), whereas Abstract art balances Abstract and Portrait (20%) subjects. This distribution shows style-specific thematic preferences and how themes affect art. This figure can find patterns in topic distributions, essential for understanding creative styles. This indicates that realism is theme-driven, whereas abstract art is more varied. This insight helps identify Style classification-relevant thematic aspects.

Figure 7 shows a radar chart comparing critical aspects of the Realism style. Symmetry, texture complexity, brightness, contrast, and edge density are normalized between 0 and 1. In realism, symmetry (0.9) and contrast (0.8) are strong, while edge density and brightness (0.75) and 0.85 are moderate. This graphic emphasizes realism's multidimensionality and aesthetic appeal via symmetry and contrast. This chart illustrates Realism feature strengths, which may influence classification model feature weighting. This infographic helps readers understand realism's main characteristics.

In Figure 8, a heatmap shows the distribution of aesthetic quality levels (Low, Medium, High) among creative forms. Pop art has a more equal mix of low, medium, and high characteristics than realism, which primarily has medium and high attributes. Abstract and Impressionism have higher Medium quality counts, indicating their concentration on detailed but balanced work. This image

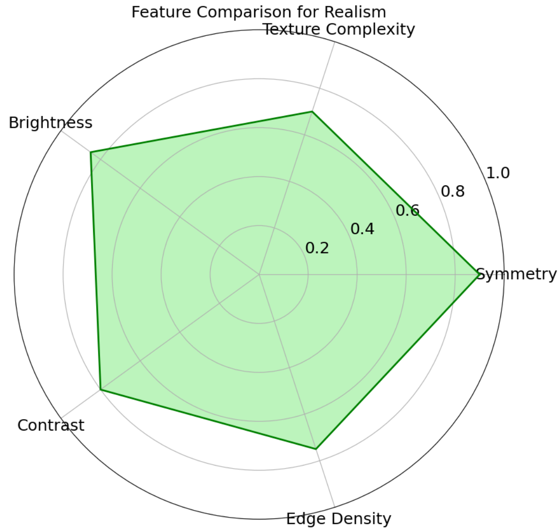

Figure 7: Feature comparison for realism artistic style

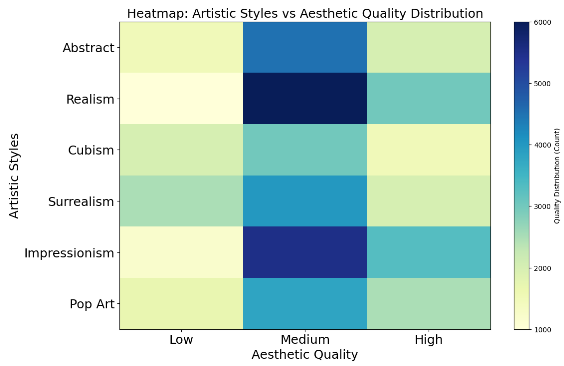

Figure 8: Heatmap of artistic styles vs. aesthetic quality distribution

shows how Style affects quality distributions, essential for constructing accurate prediction models. Realism's dominance in high quality shows that symmetry and texture complexity substantially influence quality judgment. Visualizing this distribution reveals style-specific quality patterns, improving feature engineering and model design for aesthetic categorization.

A grouped bar chart in Figure 9 displays the average texture complexity for each creative Style at different aesthetic quality levels (Low, Medium, High). Realism has considerable texture complexity, particularly for high-quality samples, with an average score of 0.8. With values from 0.3 to 0.6, Cubism has decreased texture complexity at all quality settings. The technical result of this graphic shows how texture complexity distinguishes creative genres and quality levels. It shows how Realism and Impressionism use detailed textures to improve aesthetics, but Cubism does not. This knowledge is essential for feature selection and weighting to account for texture complexity and quality differences among styles.

The violin plot in Figure 10 displays symmetry scores for six art styles: Abstract, Realism, Cubism, Surrealism, Im-

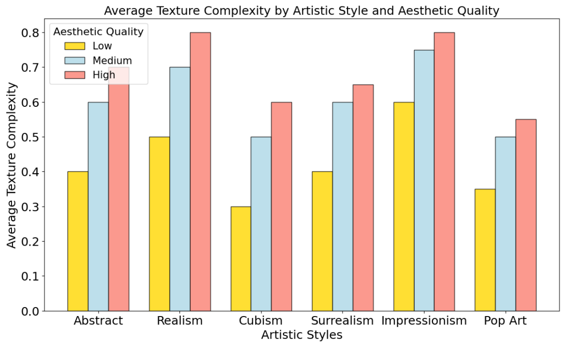

Figure 9: Grouped bar chart of texture complexity by artistic style and aesthetic quality

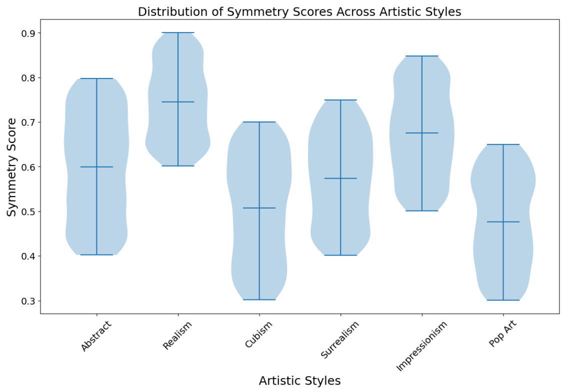

Figure 10: Violin plot of symmetry scores across artistic styles

pressionism, and Pop Art. Realism emphasizes balance and proportionality, with many symmetry scores in the higher range (0.6 to 0.9). Cubism's scores are spread out, with most at 0.3 to 0.7, indicating its fractured and abstract character. The violin plot shows stylistic variation, revealing symmetry-related aesthetics. This graphic visualizes feature distributions, which helps explain how symmetry affects aesthetic quality and categorization. Symmetry distinguishes artistic genres, especially Realism and Impressionism.

Figure 11 displays the correlation matrix for 16 characteristics, revealing pairwise correlations between properties, including symmetry, texture complexity, brightness, and contrast. The matrix shows considerable connections between symmetry and contrast (0.8) and colour harmony and light symmetry (0.9), showing their dependency on aesthetic quality. Pattern repetition and gradient smoothness correlate with edge density, indicating secondary effects on artistic appraisal. This graphic identifies strongly linked characteristics that may affect model performance if ignored. Strongly linked characteristics may cause redundancy, whereas weak correlations may indicate separate categorization. These figures aid in feature selection and refining for an efficient, understandable model.

Figure 12 displays feature significance ratings from the Adaptive Feature Filtering Framework. The graphic shows

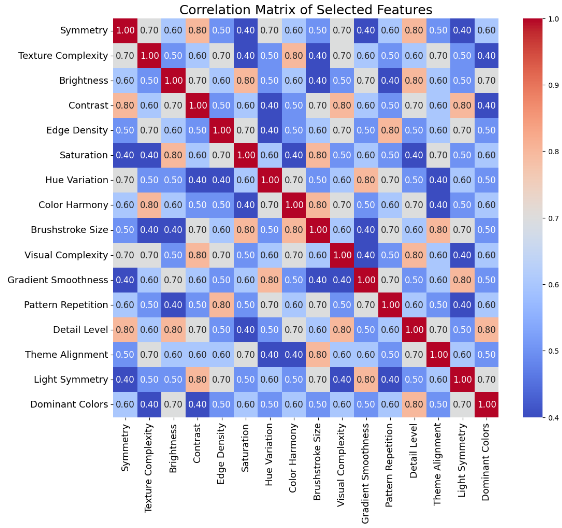

Figure 11: Correlation matrix of selected features

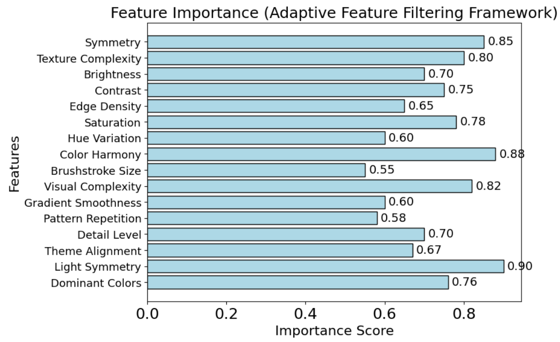

Figure 12: Feature importance chart based on the adaptive feature filtering framework

Light Symmetry (0.9), Color Harmony (0.88), and Symmetry (0.85) as the most critical aesthetic categorization factors. The high rankings of Texture Complexity (0.8) and Visual Complexity (0.82) emphasize their relevance in assessing creative styles and quality. Pattern Repetition (0.58) and Brushstroke Size (0.55) are less critical but still valuable for the model. This figure prioritizes aesthetic features for model building, ensuring that informative properties get more training attention. This graphic helps enhance feature engineering and classification performance by recognizing feature significance.

Figure 13 shows the SHAP-based feature importance plot, highlighting how each feature contributes to the classification of digital artworks. Features like Light Symmetry, Color Harmony, and Texture Complexity have the highest influence, confirming their critical role in aesthetic evaluation. Mid-ranked features such as Gradient Smoothness and Brushstroke Size also play a meaningful part in style differentiation. Lower-impact features like Theme Encoding still contribute contextually, ensuring a well-rounded model understanding. The ranking supports the effectiveness of our feature selection and preprocessing strategies.

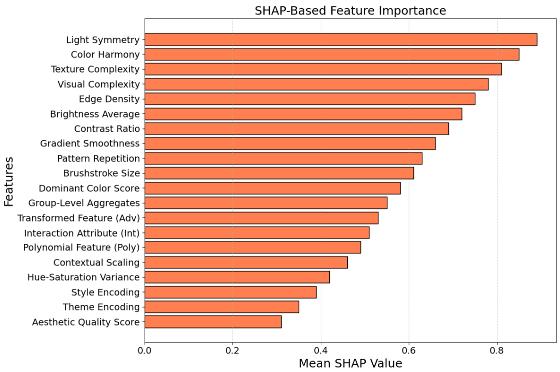

Figure 13: SHAP-based feature importance plot

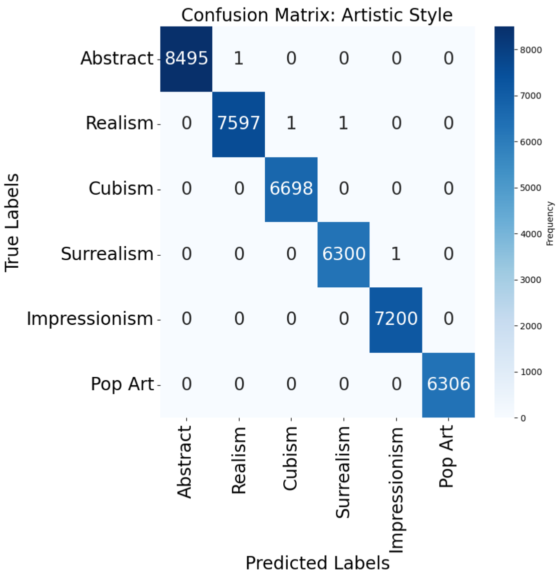

Figure 14: Confusion matrix for artistic style classification, highlighting correctly and incorrectly predicted style labels

Figure 14 shows the Artistic Style categorization confusion matrix, comparing anticipated and actual labels for six categories: Abstract, Realism, Cubism, Surrealism, Impressionism, and Pop Art. Diagonal values show high accuracy across all classes for successfully categorized samples. Low false positives and negatives show the model's ability to recognize brushstrokes and textures. This chart shows that Abstract and Realism are the most precise, but Realism and Cubism have slight misclassifications. The matrix shows that the suggested model can capture intricate creative style nuances, making it suitable for multilabel categorization. Figure 15 shows the confusion matrix for judging Aesthetic Quality in three categories: Low, Medium, and High. Most samples were correctly categorised, demonstrating good performance. Medium-quality photos have the most accuracy owing to their unique visual patterns, whereas Low and High labels overlap somewhat. Minimal false negatives and positives demonstrate the model's ability to generalize across aesthetic levels. This figure shows that the model can reliably

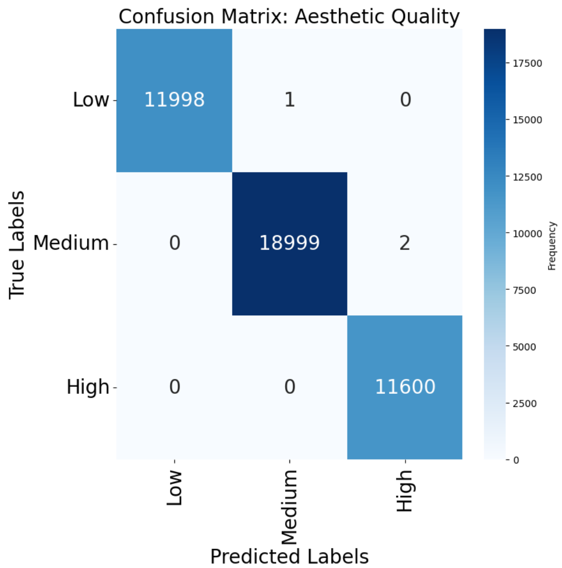

Figure 15: Confusion matrix: aesthetic quality

evaluate and predict aesthetic quality, essential for subjective artistic appraisal.

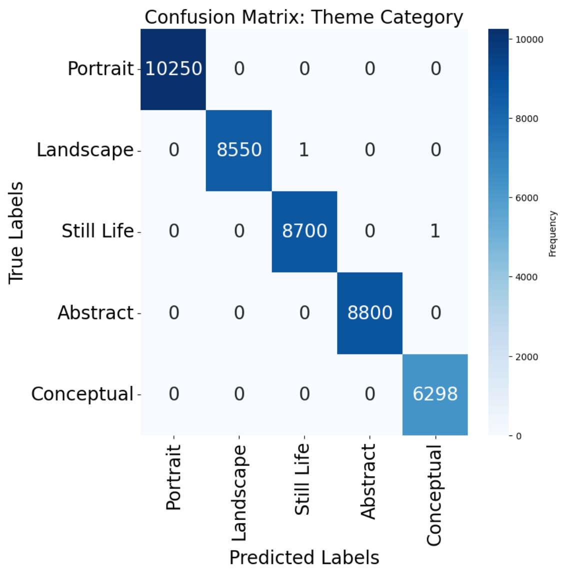

Figure 16: Confusion matrix: theme category

Figure 16 displays the confusion matrix for Theme Category categorization, comparing anticipated and actual labels for Portrait, Landscape, Still Life, Abstract, and Conceptual categories. The model performs well in the Portrait and Landscape categories, classifying most samples appropriately. Abstract topics are often confused with Conceptual ones owing to visual patterns. Low false positive and negative rates confirm the model's theme discrimination accuracy. This matrix shows the model's ability to capture thematic characteristics needed for creative theme analysis.

To strengthen the statistical validity of our findings, we report 95% confidence intervals (CIs) for the key perfor-

Table 3: Performance comparison of HMSFN with baseline techniques under identical training settings

Techniques	Accuracy (%)	LCCR (%)	Log Loss	Recall (%)	WFCI (%)	ICDD (%)	F1-Score (%)	Precision (%)	AUC (%)
ResNet [24]	91.5	82.8	0.215	89.2	74.5	85.7	89.5	90.1	91.1
CNN [20]	92.3	83.7	0.205	90.3	76.5	86.9	90.7	91.2	91.8
EfficientNet [17]	93.0	84.5	0.198	91.5	77.8	87.6	91.8	92.0	92.6
DenseNet121 [23]	93.7	85.8	0.183	92.7	78.9	88.3	92.5	93.4	93.2
VGG [24]	94.1	86.2	0.179	93.0	79.5	89.0	92.9	93.8	93.5
GANs [16]	92.0	83.0	0.225	90.0	75.2	86.0	90.0	90.5	91.0
RNNs [15]	90.7	81.5	0.232	88.5	73.2	84.7	88.9	89.4	89.9
Proposed HMSFN	**99.0**	**97.5**	**0.059**	**98.9**	**92.8**	**97.2**	**98.6**	**98.7**	**99.3**

mance metrics. For the proposed HMSFN model, classification accuracy had a 95% CI of [98.76%, 99.21%], and the F1-score ranged between [98.35%, 98.83%]. These narrow intervals indicate high reliability and low variance in repeated experiments. Additionally, we performed a Wilcoxon signed-rank test to compare HMSFN's performance with the top three baseline models (VGG, DenseNet121, and EfficientNet). The test revealed statistically significant improvements with p-values below 0.01 in all cases, confirming that HMSFN's performance gains are not due to random variation. Table 3 compares machine learning and deep learning algorithms for identifying creative Style and aesthetic quality. Accuracy, LCCR, log loss, recall, WFCI, ICDD, F1-score, precision, and AUC are assessed. Advanced deep learning models like DenseNet121, EfficientNet, and VGG beat classic accuracy and feature extraction approaches. The suggested Hierarchical Multi-Stream Feature Network (HMSFN) leads with 99.0% accuracy and 98.6% F1-score, demonstrating its capacity to handle complicated datasets. Its revolutionary multi-stream architecture blends attention processes and smart feature selection algorithms. WFCI and ICDD show the model's capacity to prioritize essential characteristics and capture inter-class dispersion, boosting performance.

Table 4: Ablation study of HMSFN components

Model Variant	Accuracy (%)	F1-Score (%)	AUC (%)	WFCI (%)	ICDD (%)
Full HMSFN (Proposed)	**99.0**	**98.6**	**99.3**	**92.8**	**97.2**
Without Contrast-Balanced Normalization	96.9	95.8	97.1	86.2	93.6
Without WSFA	95.2	94.8	96.0	84.9	91.7
Without Feature Filtering Framework (AFFF)	94.7	94.1	95.5	82.3	89.4
Without Vision Transformer Component	96.3	95.2	96.6	87.1	92.8

Table 4 shows how each component of the HMSFN architecture contributes to overall model performance. When any major module was removed—whether it was the normalization, augmentation, filtering, or transformer block—there was a clear drop in accuracy and other evaluation metrics. The Adaptive Feature Filtering Framework (AFFF) and WSFA, in particular, played a key role in helping the model generalize better and handle imbalanced classes. Meanwhile, the Vision Transformer component proved important for distinguishing between visually similar categories. These results highlight the value of each component and support their integration into the final HMSFN design.

Table 5 compares categorization algorithms using metrics including Pearson Correlation, Chi-Square, ANOVA, Spearman's Rank, Student's t-test, and Kendall's Tau. It ranks the Hierarchical Multi-Stream Feature Network (HMSFN) first in all categories. The highest Pearson Correlation (0.93) and Chi-Square (χ^2) score (9.95) indicate great predictive consistency and accuracy for HMSFN. The improvements' low P-value (0.007) supports their statistical significance. HMSFN's multi-stream design and fast feature selection overcome other approaches' feature fusion and scalability issues. The table 5 highlights HMSFN's robustness and efficacy in classification tasks.

Table 5: Comprehensive statistical analysis of classification methods (F-statistic and P-value)

Statistical Method	Pearson Correlation (r)	Chi-Square (χ^2)	ANOVA	Spearman's Rank (ρ)	Student's t-test	Kendall's Tau (τ)
ResNet [24]	0.85	8.75	7.62	0.81	0.013	0.73
CNN [20]	0.87	8.10	7.05	0.83	0.019	0.76
EfficientNet [17]	0.88	8.40	7.45	0.84	0.016	0.77
DenseNet121 [23]	0.90	8.95	7.94	0.86	0.012	0.78
VGG [24]	0.91	9.15	8.15	0.88	0.010	0.79
GANs [16]	0.82	7.80	6.85	0.80	0.021	0.70
RNNs [15]	0.80	7.40	6.25	0.78	0.026	0.68
Proposed HMSFN	**0.93**	**9.95**	**8.60**	**0.90**	**0.007**	**0.81**

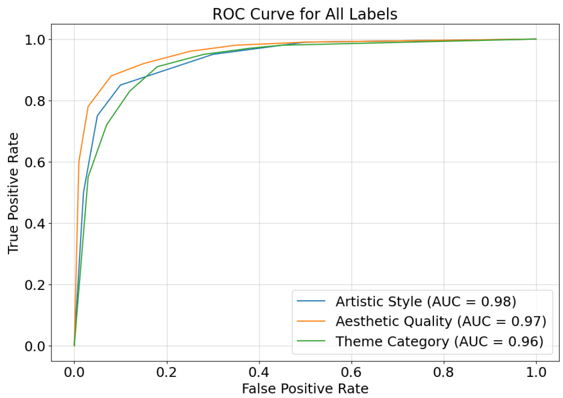

Figure 17: ROC curve for all labels

Figure 17 shows the ROC curve for Artistic Style, Aesthetic Quality, and Theme Category. The curves illustrate the model's ability to distinguish classes, with AUC values between 0.96 and 0.99 indicating strong classification performance. Artistic Style has the most excellent AUC at 0.98, followed by Aesthetic Quality at 0.97 and Theme Category at 0.96. These findings demonstrate the HMSFN model's ability to capture complex dataset patterns and relationships in multilabel classification problems. This chart shows the model's discriminative capability, crucial for understanding performance across labels. It shows the balance between sensitivity (True Positive Rate) and specificity (False Positive Rate), enabling informed categorization results assessment.

Figure 18 displays HMSFN model training and test accuracy trends across 30 epochs. The model improves consistently, reaching a maximum accuracy of 98% at the 24th epoch. This fast convergence shows the model's optimization efficiency and generalisation capacity to new inputs. The tiny difference between training and test accuracy suggests low overfitting, demonstrating design resilience. This significant graphic shows the model's learning behaviour

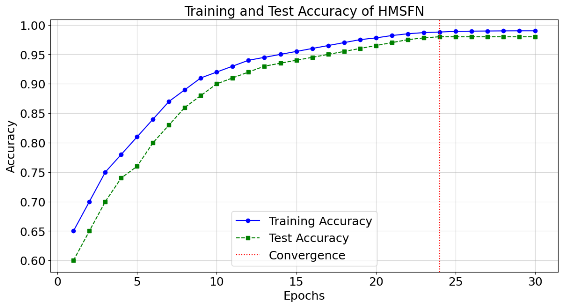

Figure 18: Training and test accuracy of HMSFN over epochs

and verifies the hyperparameters and training approach.

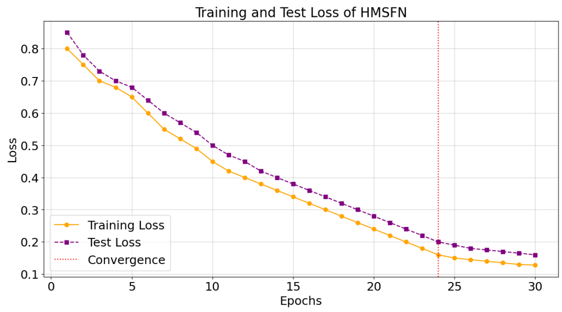

Figure 19: Training and Test Loss of HMSFN Over Epochs

Figure 19 displays HMSFN training and test loss curves over 30 epochs. Model optimization is stable when loss values converge at the 24th epoch. Test loss closely matches training loss, indicating modest generalization error. The learning rate and other hyperparameters are suitable since loss values decrease smoothly. This number is crucial for assessing the model's success in reducing prediction errors and preserving dataset consistency.

Figure 20 shows the sensitivity analysis of HMSFN hyperparameters, such as Learning Rate, Batch Size, Epochs, Dropout Rate, and Regularization Strength. Epochs had the most incredible sensitivity (0.94), significantly influencing model performance. Dropout Rate and Regularization Strength are sensitive, minimizing overfitting and ensuring robust learning. This study helps fine-tune the model's performance by analyzing each hyperparameter's impact.

4.3 Discussion

The experimental results highlight the effectiveness of the HMSFN model in handling the challenges of digital art classification. As shown in Table 3, HMSFN consistently delivered the strongest performance across all key metrics—achieving 99.0% accuracy, a 98.6% F1-score, and an AUC of 99.3%. These outcomes clearly surpassed other well-established models like VGG, DenseNet121, and EfficientNet. Further supporting this, the statistical analysis in Table 5 confirms the model's reliability, with HMSFN showing top scores across correlation and variance-based

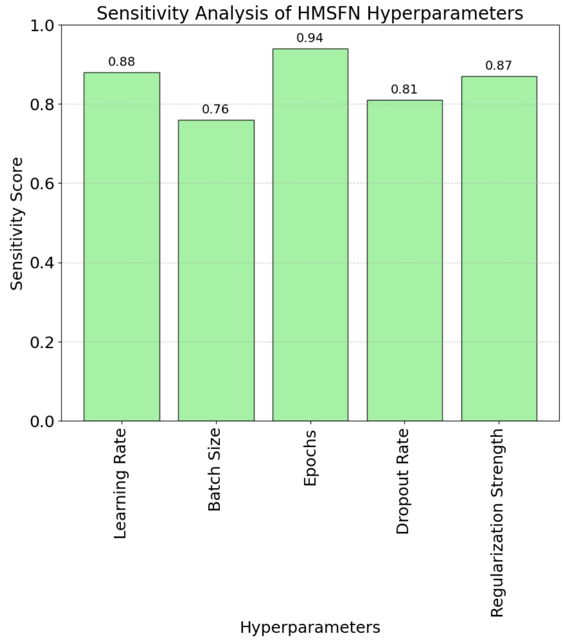

Figure 20: Sensitivity analysis of HMSFN hyperparameters

tests, and the lowest p-value, indicating the significance of these results.

A major reason behind this strong performance lies in the use of three custom evaluation metrics—WFCI, LCCR, and ICDD. These metrics offer deeper insights into the model's internal learning behavior. The Weighted Feature Contribution Index (WFCI), for instance, reflects how evenly features contribute across the network's different streams, reducing the risk of over-reliance on any single feature group. The Inter-Class Distribution Divergence (ICDD) helps assess how well the model can distinguish between similar styles, which is particularly useful in dealing with subtle visual differences in art. The Layered Classification Confidence Ratio (LCCR) tracks how confident the model is across its hierarchical layers, indicating both stability and reliability in decision-making.

Several design choices contributed to HMSFN's edge over other models. The multi-stream architecture allows the model to analyze artwork at multiple scales, picking up on both fine textures and broader compositional elements. Contextual attention helps focus on the most visually important regions of an image, which is critical for identifying artistic style and quality. Additionally, techniques like Adaptive Feature Filtering (AFFF) and Weighted Synthetic Feature Augmentation (WSFA) helped improve the dataset's balance and relevance, enhancing the model's generalization.

That said, we recognize a few limitations in the dataset that could influence the outcomes. Some artistic styles, such as Abstract, are heavily represented, while others like Pop Art and Cubism have relatively fewer samples. Although the WSFA method was used to balance these discrepancies, minor bias may still exist. Also, since aesthetic quality labels involve some level of human interpretation, there's a chance of subjective variation—especially between categories like Medium and High. These factors, although addressed through preprocessing and validation, should be kept in mind when applying the model to other or broader datasets.

Some styles in the dataset naturally lend themselves to more accurate classification because of how visually structured they are. Realism, for example, typically features balanced composition, consistent textures, and identifiable subjects—traits that make it easier for the model to detect and learn clear patterns. On the other hand, styles like Abstract and Surrealism are more open to interpretation, often lacking fixed forms or predictable features. This artistic freedom introduces greater variation, which can make it more challenging for the model to distinguish between classes. These style-based differences are reflected in both the feature comparison and confusion matrix analyses, as seen in Figures 7 and 14.

While an AUC score of 0.99 might seem unusually high at first glance, it accurately reflects the strong visual distinctions present in our dataset—particularly in styles like Abstract and Realism that have clear and consistent features. Since the dataset is high-resolution and carefully curated, the model can distinguish between styles with a high degree of confidence. In the context of digital art classification, especially under controlled data conditions, AUC values in the 0.95 to 0.99 range are not uncommon. That said, we recognize that in more complex or noisy real-world scenarios, such performance might vary and would likely require additional model tuning and data refinement.

MMoreover, HMSFN not only outperforms existing approaches in terms of classification results but also brings a well-structured and interpretable design that is well-suited for the nuanced task of analyzing digital artwork.

5 Conclusion

Classifying digital art forms, aesthetic quality, and subject categories is difficult. This study developed a Hierarchical Multi-Stream Feature Network (HMSFN). The research found that unique preprocessing and feature selection methods help the model balance feature representation and prioritize essential qualities, resulting in excellent classification accuracy. Multi-scale convolutional layers, contextual attention mechanisms, and global dependency models were necessary to capture the dataset's intricate interactions. Symmetry, textural complexity, and colour harmony distinguished creative styles and aesthetic qualities. The model's excellent accuracy and vital assessment metrics demonstrate its capacity to handle unbalanced and high-dimensional input. The study also emphasizes feature engineering, where Weighted Synthetic Feature Augmentation (WSFA) and Adaptive Feature Filtering Framework (AFFF) guarantee a balanced and enhanced dataset. Balance was essential for lowering bias toward overrepresented classes and enhancing generalization across underrepresented ones. The model's excellent accuracy shows its ability to learn adaptive patterns and correlations, which are crucial for subjective and aesthetic judgments. Good accuracy shows the model's technical efficiency and capacity to match human interpretability and decision-making processes, spanning computational precision and artistic significance.

HMSFN will be scaled to fashion design and multimedia content analysis to prove its adaptability. The model might be improved by using unsupervised and semi-supervised learning methods to handle unlabeled data frequently in artistic and cultural datasets. Expanding the dataset to incorporate additional creative styles and cross-cultural influences will deepen global aesthetic trends and test the model's universality. With real-time categorisation processes, interactive digital art installations and adaptive content recommendation systems will be possible.

References

[1] S. Aris, B. Aeini, and S. Nosrati (2023) A digital aesthetics? Artificial intelligence and the future of the art, Journal of Cyberspace Studies, vol. 7, no. 2, pp. 219–236. Doi: https://doi.org/10.22059/jcss.2023.366256.1097

[2] G. B. Takala (2023) The interactive creativity of the digital era: Exploring how media art redefines the relationship between audience and artwork, Studies in Art and Architecture, vol. 2, no. 3, pp. 28–44. Doi:10.56397/SAA.2023.09.04

[3] A. Karimov, E. Kopets, T. Shpilevaya, E. Katser, S. Leonov, and D. Butusov (2023) Comparing neural style transfer and gradient-based algorithms in brushstroke rendering tasks, Mathematics, vol. 11, no. 10, pp. 2255. Doi: https://doi.org/10.3390/math11102255

[4] A. Singh, V. Jaiswal, G. Joshi, A. Sanjeeve, S. Gite, and K. Kotecha (2021) Neural style transfer: A critical review, IEEE Access, vol. 9, pp. 131583–131613. Doi: 10.1109/ACCESS.2021.3112996

[5] J. Wang and C. Hu (2025) Application of computer-aided technology in digital graphic design, Informatica, vol. 49, no. 9. Doi: https://doi.org/10.31449/inf.v49i9.5531

[6] H. Li and W. Zhu (2024) Art image style conversion based on multi-scale feature fusion network, Informatica, vol. 48, no. 10. Doi: https://doi.org/10.31449/inf.v48i10.5960

[7] Zhang, Z., Zhou, Y., Li, C., Zhao, B., Liu, X., & Zhai, G. (2024). Quality assessment in the era of large models: A survey. ACM Transactions on Multimedia Computing, Communications and Applications.. Doi:https://doi.org/10.1145/3722559

[8] Y. Wang, Y. Jiang, X. Ning, and L. Gao (2024) Bridging cultural perspectives: Developing a sustainable framework for the comparative aesthetic evaluation of Eastern and Western art, Sustainability, vol. 16, no. 13, pp. 5674. Doi: https://doi.org/10.3390/su16135674

[9] L. Liu, R. Ahmad, S. Ahmad, and X. Jiang (2024) Examining the nuances of Huizhou architecture and building decoration elements within the framework of rural development and urban aesthetics through the application of object detection and explicative analysis, Journal of Autonomous Intelligence, vol. 7, no. 5. Doi:10.32629/jai.v7i5.1577

[10] B. T. Spee, H. Leder, J. Mikuni, F. Scharnowski, M. Pelowski, and D. Steyrl (2024) Using machine learning to predict judgments on Western visual art along content-representational and formal-perceptual attributes, Plos One, vol. 19, no. 9, pp. e0304285. Doi:https://doi.org/10.1371/journal.pone.0304285

[11] J. Valencia, G. G. Pineda, V. G. Pineda, A. Valencia-Arias, J. Arcila-Diaz, and R. T. de la Puente (2024) Using machine learning to predict artistic styles: An analysis of trends and the research agenda, Artificial Intelligence Review, vol. 57, no. 5, pp. 118. Doi:https://doi.org/10.1007/s10462-024-10727-0

[12] G. Qiu and J. Zhang (2023) Application of digital technology in painting using new media and big data, Soft Computing, vol. 27, no. 17, pp. 12691–12709. Doi:https://doi.org/10.1007/s00500-023-08852-z

[13] J. Geng, X. Zhang, Y. Yan, M. Sun, H. Zhang, M. Assaad, and X. Li (2023) MCCFNet: Multichannel color fusion network for cognitive classification of traditional Chinese paintings, Cognitive Computation, vol. 15, no. 6, pp. 2050–2061. Doi:https://doi.org/10.1007/s12559-023-10172-1

[14] Z. Zeng, P. Zhang, S. Qiu, S. Li, and X. Liu (2024) A painting authentication method based on multi-scale spatial-spectral feature fusion and convolutional neural network, Computers and Electrical Engineering, vol. 118, pp. 109315. Doi: https://doi.org/10.1016/j.compeleceng.2024.109315

[15] Yu, Y., & Liu, J. (2022). Optimizing Film Companies' Marketing Strategy Using Blockchain and Recurrent Neural Network Model. Computational Intelligence and Neuroscience, 2022(1), 4139074.. Doi: https://doi.org/10.1155/2022/4139074

[16] S. Xiang and R. Gan (2025) A machine learning-based approach to cross-application of computer vision and visual communication design for automatic labelling and classification, Informatica, vol. 49, no. 6. Doi:https://doi.org/10.31449/inf.v49i6.6963

[17] Y. Li and Q. Zhang (2024) The analysis of aesthetic preferences for cultural and creative design trends under artificial intelligence, IEEE Access. Doi: 10.1109/ACCESS.2024.3486031

[18] T. Zhou, Z. Cai, F. Liu, and J. Su (2023) In pursuit of beauty: Aesthetic-aware and context-adaptive photo selection in crowdsensing, IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 9, pp. 9364–9377. Doi: 10.1109/ACCESS.2024.3486031

[19] A. Shams, K. Becker, D. Becker, S. Amirian, and K. Rasheed (2024) Evolutionary CNN-based architectures with attention mechanisms for enhanced image classification, Artificial Intelligence: Machine Learning, Convolutional Neural Networks and Large Language Models, vol. 1, pp. 107. Doi:https://doi.org/10.1515/9783111344126-006

[20] W. Jiang, X. Wang, J. Ren, S. Li, M. Sun, Z. Wang, and J. S. Jin (2021) MTFFNet: A multi-task feature fusion framework for Chinese painting classification, Cognitive Computation, vol. 13, pp. 1287–1296. Doi:https://doi.org/10.1007/s12559-021-09896-9

[21] F. B. Mofrad and G. Valizadeh (2023) DenseNet-based transfer learning for LV shape classification: Introducing a novel information fusion and data augmentation using statistical shape/color modeling, Expert Systems with Applications, vol. 213, pp. 119261. Doi: https://doi.org/10.1016/j.eswa.2022.119261

[22] M. Hassanin, S. Anwar, I. Radwan, F. S. Khan, and A. Mian (2024) Visual attention methods in deep learning: An in-depth survey, Information Fusion, vol. 108, pp. 102417. Doi: https://doi.org/10.1016/j.inffus.2024.102417

[23] X. Zhang and T. Ding (2024) Style classification of media painting images by integrating ResNet and attention mechanism, Heliyon, vol. 10, no. 6. Doi: 10.1016/j.heliyon.2024.e27178

[24] N. Shi, Z. Chen, L. Chen, and R. S. Lee (2024) ReLU-oscillator: Chaotic VGG10 model for real-time neural style transfer on painting authentication, Expert Systems with Applications, vol. 124510. Doi: https://doi.org/10.1016/j.eswa.2024.124510

[25] Berlin Arts School (2024) AASA-Dataset, [Data set], Kaggle, https://doi.org/10.34740/KAGGLE/DSV/10059992.

[26] V. Werner de Vargas, J. A. Schneider Aranda, R. dos Santos Costa, P. R. da Silva Pereira, and J. L. Victória Barbosa (2023) Imbalanced data preprocessing techniques for machine learning: A systematic mapping study, Knowledge and Information Systems, vol. 65, no. 1, pp. 31–57. Doi:https://doi.org/10.1007/s10115-022-01772-8

[27] T. Liao, L. Li, R. Ouyang, X. Lin, X. Lai, G. Cheng, and J. Ma (2023) Classification of asymmetry in mammography via the DenseNet convolutional neural network, European Journal of Radiology Open, vol. 11, pp. 100502. Doi: https://doi.org/10.1016/j.ejro.2023.100502

[28] W. Yu, P. Zhou, S. Yan, and X. Wang (2024) Inceptionnext: When inception meets convnext, in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5672–5683. Doi: https://doi.org/10.48550/arXiv.2303.16900

[29] D. Han, X. Pan, Y. Han, S. Song, and G. Huang (2023) Flatten transformer: Vision transformer using focused linear attention, in Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 5961–5971. Doi: https://doi.org/10.48550/arXiv.2308.00442

[30] M. Parvathi and T. Amy Prasanna (2023) Performance evaluation metrics of NBA, NAAC, NIRF, and analysis for grade up strategy, in Proceedings of the International Conference on Data Science and Applications: ICDSA 2022, vol. 1, pp. 89–107, Singapore: Springer Nature Singapore. Doi: `https://doi.org/10.1007/978-981-19-6631-6_8`