SCIENTIFIC OASIS

Decision Making: Applications in Management and Engineering

Journal homepage: www.dmame-journal.org
ISSN: 2560-6018, eISSN: 2620-0104

A ResNet-Enhanced Framework for Modelling Aesthetic Preferences in User-Centred Decision Systems

Jiadi Liu[1], Yixin Zhang[2,*], Shuhui Liu[3]

1 BA, School of Architecture and Design, Beijing Jiaotong University, Beijing 100044, China. Email:22724017@bjtu.edu.cn
2 BA, Art Institute, Lanzhou University, Lanzhou, China, 730,000. Email:320220944270@lzu.edu.cn
3 BA, School of Mechanical Engineering, Dongguan University of Technology, Dongguan China, 523808. Email:shuihuiliu0211@163.com

ARTICLE INFO

Article history:
Received 25 April 2024
Received in revised form 30 May 2024
Accepted 7 June 2024
Available online 29 June 2024

Keywords:
Aesthetic Preference Modelling, Multimodal Interaction, Deep Learning, Multimedia Analysis, Decision Support Systems

ABSTRACT

Modelling user aesthetic preferences is essential in contemporary multimodal interaction systems, as it enhances content relevance, user engagement, and decision-making efficacy. This study introduces a deep learning-based framework that integrates visual feature extraction with user interaction data to estimate the aesthetic quality of multimedia content, using images and videos as illustrative cases. The framework utilises statistical image analysis as input to deep convolutional architectures, specifically Residual Neural Networks (ResNets), to compute key aesthetic attributes such as colour harmony, lightness, and visual complexity. Additionally, user engagement indicators, including likes and collections, are incorporated alongside these visual features to infer patterns of aesthetic preference. Given the inherent subjectivity and variability in user data, the framework applies annotation normalisation and layer-freezing strategies to enhance model generalisation and training efficiency. The proposed system demonstrates robust capabilities in aesthetic scoring and preference modelling, facilitating informed content presentation and decision support across various digital environments. This work contributes to the advancement of intelligent, user-aware systems within the domains of multimedia interaction and human-computer interface design.

1. Introduction

The utilisation of interactive systems has progressed rapidly during the era of digital transformation, responding to the evolving requirements of users [1]. These systems can generally be categorised into two main types, particularly multimodal interaction platforms that incorporate diverse input and output modalities such as voice, touch, gesture, and visual interfaces. This categorisation is due to their adaptability and user-centred orientation [2]. These platforms are designed to emulate natural human communication patterns while also enhancing system usability [3]. As the user base becomes increasingly diverse, expectations for how such systems should function, appear, and respond continue to grow. Among the influencing factors, aesthetic preference stands out as both a critical and underexplored aspect. This qualitative dimension significantly

* *Corresponding author.*
E-mail address: 320220944270@lzu.edu.cn

https://doi.org/10.31181/dmame7120241491

impacts user engagement, decision-making, and overall satisfaction with interactive systems [4]. Integrating these factors into system models has the potential to elevate the quality of user interaction and the effectiveness of decision-making processes facilitated by these systems.

Historically, aesthetics in human-computer interaction (HCI) have been confined to visual design considerations such as colour palettes, layout configurations, and typography. Although these elements remain relevant, the concept of user experience has since expanded to include dynamic interface behaviour, personalisation features, cultural contexts, and users' emotional responses to system feedback [5]. When users engage with complex systems—particularly those designed to support decision-making—their aesthetic judgements subtly influence trust in the system, perceived intuitiveness, and frequency of use [6]. These influences are particularly pronounced in decision support contexts, where cognitive load and emotional states are crucial. Thus, integrating models that recognise and respond to aesthetic preferences can result in systems that are not only more effective but also more user-friendly [7].

Moreover, contemporary systems often approach aesthetic preference modelling in a multimodal manner, presenting both opportunities and complexities [8]. Multimodal interactions comprise layered exchanges—such as the synchronisation of auditory and visual cues or the coordination of haptic feedback with gesture inputs—which together create a rich, dynamic experience. In these contexts, aesthetics extend beyond visual design into multisensory domains [9]. For instance, a user may perceive a system as aesthetically superior when auditory feedback aligns seamlessly with visual animations, or when tactile responses to gestures feel natural and responsive [10]. The broad spectrum of device interactions within a multimodal environment, combined with the inherently ecological nature of user experiences, indicates that design approaches must transcend traditional graphical user interface paradigms. Instead, effective design must incorporate behavioural data, context awareness, machine learning methodologies, and human-centred evaluations to tailor experiences across different modalities [11].

With the growing adoption of decision support systems (DSS) across sectors such as healthcare, finance, urban planning, and education, the personalisation of user interactions has become increasingly vital [12]. These systems typically operate in environments characterised by high complexity and dense data, where users must make swift yet informed decisions with a high degree of confidence [13]. Aesthetic modelling holds promise in enhancing decision quality by mitigating cognitive friction, aligning system behaviour with user expectations, and fostering trust in the system [14]. For example, a DSS interface supporting a resident doctor can present relevant information in a manner that is visually coherent, emotionally reassuring, and cognitively manageable, thereby improving the interaction experience [15]. Embedding aesthetic preference modelling within such systems brings us closer to developing more empathetic, capable, and efficient tools for decision-making support [16].

This research outlines a framework for modelling user aesthetic preferences in the context of multimodal interaction systems, particularly within decision support applications. The core components discussed include the generation of aesthetic feedback via various sensing modalities, strategies for preference analysis, and the selection of suitable algorithms. The study bridges the operational mechanisms of interactive systems with user interface experience, leveraging insights from human-computer interaction, cognitive psychology, artificial intelligence, and design theory. Ultimately, the research aspires to enhance human decision-making capabilities by uncovering how aesthetic considerations shape user engagement with technological systems.

2. Related Works

Understanding and modelling users' aesthetic preferences in interaction systems that utilise multiple input modalities is essential for enhancing user experience, particularly in decision support applications. As human-computer interaction becomes increasingly complex and personalised, both academic and industrial research have begun to emphasise the importance of interpreting and adapting to users' subjective preferences across visual, auditory, and tactile modalities [17]. Advancements in machine learning, affective computing, and multimodal data processing have made it possible to implement more sophisticated methods for capturing such preferences, thereby enabling systems to dynamically adjust the content, appearance, and flow of interaction interfaces [18]. Techniques such as convolutional neural networks, reinforcement learning, and feedback loops derived from user studies have proven effective in facilitating real-time adaptation and improving decision-making accuracy.

This literature review examines significant contributions in this domain, focusing on how computational models, adaptive frameworks, and sensory integration techniques have been applied to the modelling of aesthetic preferences. It also highlights existing technical limitations and identifies research gaps that present opportunities for further investigation. Maroto-Gómez et al. [19] proposed an adaptive decision-making system that forecasts user preferences to improve human-robot interactive communication. Their approach involves the application of machine learning algorithms to analyse user behaviour and subsequently adapt the robot's responses dynamically throughout the interaction. The system possesses the capability to progressively align with users' interaction patterns, thereby increasing engagement and minimising frustration. A key strength of this system is its capacity for real-time personalisation, rendering it highly applicable in the field of assistive robotics. Nonetheless, a notable limitation lies in the computational burden associated with collecting reliable user data over extended periods, which can complicate system scalability in broader applications. The validation of the research gap is presented in Table 1.

Table 1
Research Gap Validation

Author(s)	Techniques Involved	Advantages	Disadvantages
Peng and Wang [20]	Adaptive Decision-Making, Preference Prediction	Personalized, Real-Time Interaction	High Computational Load, Data Dependency
Dauber-Decker et al. [15]	UIED Segmentation, GUI Evaluation	Fast, Objective UI Assessment	Ignores Subjective/Emotional Factors
Nie et al. [21]	Fuzzy Logic in Multimodal Systems	Handles Uncertainty, Dynamic Adaptability	Rule-Based Limits Scalability
Zhang et al. [22]	Decision Support Via Multi-Criteria in MATLAB	Structured, User-Friendly, Efficient	Expert Bias, Less Flexible
Yuan et al. [23]	AHP-Based Multimodal Evaluation	Systematic, Quantitative Comparison	Subjective Input, Setup Overhead

Kyelem et al. [24] introduced an automated tool for evaluating the aesthetics of graphical user interfaces (GUIs) using the UIED segmentation algorithm. This system applies computer vision methodologies to segment interface components and assess their stylistic quality based on the structured visual layout. A notable advantage of this approach lies in its objectivity and automation, as it removes dependence on subjective human evaluation. Consequently, design assessments can be performed more rapidly and consistently compared to traditional, subjective methods, thus facilitating quicker prototyping through early feedback during the design phase. However, a key limitation of the method is its inability to interpret subtle user preferences or emotional responses to visual elements, which play a significant role in user experience (UX) design.

Zhang et al. [22] explored multimodal interaction in dynamic contexts through the application of fuzzy logic theory to perception and decision-making processes. Their system integrates decision-making functions with fuzzy logic to allow adaptation in response to fluctuating environmental conditions and user states. Although the approach is grounded in mathematical certainty for discrete datasets, its principal strength is the ability to manage uncertainty and imprecise data—both of which are characteristic of real-world applications. In environments with complexity or ambiguity, this approach contributes to more robust interaction performance. Nevertheless, a core limitation is that users must manually define rules and membership functions, making the system reliant on domain-specific configurations, which hinders scalability and general applicability.

Özdiler Çopur et al. [25] created a DSS aimed at assisting with strategic purchasing decisions for dental implants, implemented via a MATLAB-based GUI. The system employs a multi-criteria decision-making model that leverages expert input to assess implant alternatives across criteria such as cost, material type, and clinical success rates. Its strength lies in the transparency and structure of the decision-making process, which enhances procurement accuracy and operational efficiency. Additionally, the visual interface improves accessibility for users without technical backgrounds. However, the method's reliance on expert-defined weights and evaluation criteria introduces potential bias and limits the system's ability to adapt across varying clinical scenarios. Peng and Wang [20] presented a User-Device Interaction model for analysing multimodal interaction using the Analytic Hierarchy Process (AHP). This system facilitates a quantitative comparison of interaction modes—such as voice, touch, and gesture—by decomposing user preference evaluations into a structured hierarchical framework. A key benefit of this approach is its systematic nature, which supports consistent evaluations across metrics like responsiveness, accuracy, and user comfort. However, the process can be time-consuming to establish, and the outcomes rely heavily on subjective user judgments, which may vary significantly and lead to inconsistencies.

From the literature reviewed, several limitations are evident, including dependence on manually defined rules, subjective input from users, and restricted scalability in dynamic interaction environments. To address these challenges, this study proposes the use of ResNet as a core model for capturing and analysing user aesthetic preferences in multimodal interaction systems. ResNet, a deep convolutional neural network architecture, addresses the problem of degradation in deeper networks through residual learning. This allows the model to extract meaningful features from complex, high-dimensional datasets. One of ResNet's key advantages is its capacity to learn hierarchical features from a wide array of input modalities, including visual signals, interface behaviours, and user feedback. This represents a substantial improvement over traditional systems that rely on rule-based or manually tuned configurations. Furthermore, ResNet supports large-scale deployment and real-time adaptability, making it well-suited to dynamic decision support applications that require high accuracy and responsiveness. The proposed system thus aims to enhance personalisation, reduce bias, and improve user satisfaction in intelligent multimodal environments through the integration of ResNet technology.

3. Proposed System Model

Research in computational aesthetics estimates aesthetic judgements through methods linked to subjective statistical image characteristics and perceptual cues. Investigations have primarily focused on two key aspects of effective aesthetic validation in multimedia presentation: the selection of appropriate techniques (such as machine learning models) and the construction of suitable datasets (including the acquisition of multimedia content or aesthetic stimuli). By identifying calculable parameters of visual features, suitable methods for deriving aesthetic judgements, and optimal training strategies, researchers have established potential directions and practical applications for aesthetic modelling [26]. GUI features are often evaluated using machine learning, statistical correlation, and quantitative analysis. Given the significance of assessing websites based on measurable characteristics, numerous related metrics have emerged. In recent developments, growing attention has been directed toward the application of machine learning techniques for the aesthetic evaluation of human facial imagery [27].

With the proliferation of social media usage, individuals increasingly generate and engage with substantial volumes of user-generated content, reflecting their personal preferences and aesthetic sensibilities [28]. Among the various forms of multimedia, images represent a primary modality for user expression, thus serving as a valuable source of data on aesthetic preferences. In the present model, ResNet is employed as a preferred framework for computational aesthetics. GUI images are collected and resized to a uniform dimension—without cropping—to accommodate architectural requirements [29]. Initially, the input images undergo processing via the first layer of convolutional filters. The study leverages four distinct deep learning architectures to process the image inputs. The resulting output produces a probability distribution over a set of 'like' and 'dislike' labels, which can serve as indicators of the user’s aesthetic preferences. The proposed methodology follows a two-stage pipeline: the first stage involves feature extraction through ResNet, and the second stage pertains to the development of the aesthetic computation model. To evaluate GUI-related content based on user interactions such as 'likes' and image collections, the system identifies the most suitable structure and compares it with established aesthetic assessment methods to optimise aesthetic quality. The model is further refined for use in decision-making contexts. A visual representation of the complete system architecture is provided in Figure 1.

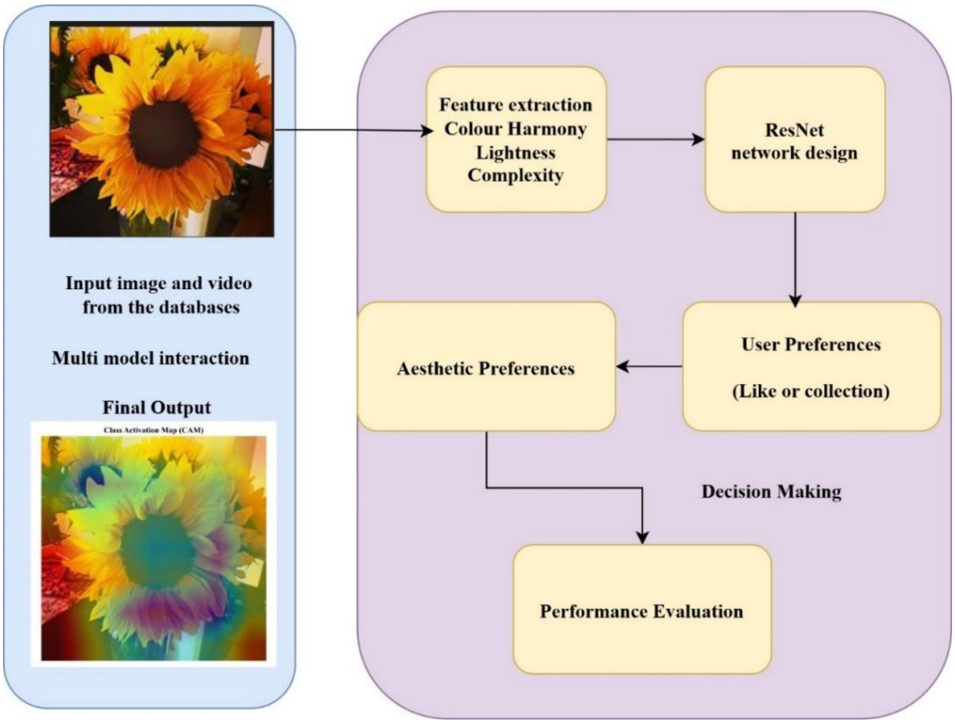

Figure 1: Block Diagram of the Proposed Model

3.1 GUI Image Data Collection

Given the extensive volume of GUI designs already produced by digital image creators, the compiled database of images and videos is sufficiently comprehensive to enable effective aesthetic evaluation of design potential. Visual materials were sourced from publicly available datasets consisting of design images and videos. A real-world index was constructed utilising user-generated ground truth labels such as collections and likes, which serve as proxies to estimate the aesthetic value of batches of GUI images and videos. This approach is founded on the premise that visual appeal is likely the principal factor influencing the frequency of collections and likes. Furthermore, this index can be employed to predict aesthetic trends based on the number of collections and likes received. In total, 38,423 samples were gathered. For consistency, all input videos and images were standardised to a fixed size without cropping and were collected at a resolution of 72 dpi. The distribution of samples from the database, plotted with likes and collection counts, uses images and videos as the independent variable (abscissa) and the corresponding number of likes as the dependent variable (ordinate). Several representative samples are depicted to illustrate this relationship. To address the challenge posed by data dispersion, annotation normalisation was performed. This normalisation process was applied specifically to selected annotations to improve data consistency and reliability.

3.2 Feature Extraction

The proposed approach for general aesthetic phase computation integrates evaluation criteria such as colour harmony, brightness, and complexity. The underlying interior design principles include lightness, which quantifies image brightness; colour harmony, reflecting the effective utilisation of multiple colours within an image; and complexity, which represents the visual intensity of design elements within the spatial composition. This study employs image processing techniques to facilitate automatic feature extraction. Colour harmony assessment is conducted through the analysis of colour histograms. Brightness measurement is derived by quantifying the proportions of dark and light regions within the image. Complexity analysis involves enumerating objects present in the image, supported by precise edge detection algorithms for parameter identification. The overall aesthetic score is calculated as a weighted average of these extracted features. The ResNet model operates in conjunction with colour preference data to generate the aesthetic score, which is subsequently used to determine the aesthetic preference [30]. The method operates according to the following description.

3.2.1 Colour Harmony

The degree of colour harmony is assessed by analysing three colour attributes: intensity, saturation, and hue structure. The image conversion process marked the initial phase in developing the proposed framework. Each image within the database retained colour harmony levels scaled between 0 and 100. An overall sense of harmony in colour composition arises when images are combined, as this process activates relevant statistical colour combinations that are subsequently compared against established colour harmony criteria. The procedures for similarity measurement and similarity evaluation are conducted sequentially.

3.2.2 Lightness

The assessment of image brightness involves applying a generalised formula to quantify lightness metrics. The algorithm assigns a value ranging from 0 to 10, where 0 represents low brightness and 10 corresponds to high brightness. Utilising the Python Imaging Library, the programme analyses each opened image to determine its brightness level.

3.2.3 Complexity

Visual complexity is defined as the measure of all elements present within an interior space. The visual potential and resulting visual weight increase proportionally with the number of items contained in space. The quantity of interior elements directly determines the amount of visual information requiring processing. Elevated visual complexity can negatively impact design appeal and overall compositional balance. The complexity is quantified using an intelligent algorithm. This involves a multi-stage process of advanced edge detection technology, which actively extracts image edges. The method identifies image boundaries as part of this procedure. Detection of prominent and distinct edges is performed by this technique, which serves a fundamental role in various image processing tasks.

3.3 ResNet Network Design

ResNet addresses the issue of gradient vanishing by enhancing gradient flow through the introduction of shortcut connections that provide an alternative pathway. Unlike conventional feedforward networks, these shortcut connections model the residual function, facilitating more efficient optimisation compared to direct mapping. A deeper network is expected not to exhibit higher error rates than its shallower counterpart, as additional layers can be utilised to refine the mapping. Residual mapping involves multiple stacked layers that extend beyond basic layers, enabling the network to approximate complex underlying functions rather than only specific mappings [21]. The nonlinear layers construct various transformations that collectively approximate the target function denoted as H(X).

$$F(X) = H(X) - X \qquad (1)$$
$$H(X) = F(X) + X \qquad (2)$$

In this context, H(X) denotes the original mapping, while F(X) represents the residual function, with the architecture designed to train the residual F(X). The shortcut connections bypass intermediate layers, allowing the residual function to be learned more effectively. The residual function is computed through simple vector addition, which facilitates backpropagation during residual training. Importantly, the addition of these residual mappings does not introduce any new parameters into the system. Although ResNet architecture may vary in the number of layers, their operation remains consistent with this fundamental principle.

3.3.1 Layer Freezing

Convolutional Neural Network (CNN) models are utilised through two principal approaches to leverage their learned capabilities. The first approach preserves the pre-trained CNN architecture intact, employing its feature extraction capacity to assist in training and validation. The second approach involves optimising CNN by modifying existing design components and parameters to enhance performance. The introduction of new learnable parameters allows the architecture to retain specific historical data relevant to the task. While the initial approach reduces the computational expense associated with training the entire deep network, it often entails greater complexity compared to the latter strategy. Deep architectures tend to overfit when trained on small datasets, resulting in unsatisfactory performance on test data; consequently, the number of feature detectors activated per epoch should be reduced to mitigate overfitting effects [31]. Moreover, deeper layers demand the most computational resources, even when hyperparameters are minimised. Standard hyperparameters are employed across CNNs to reduce unnecessary computational burden, often by preserving trained feature extractor models from prior training phases. Learning is primarily concentrated in the convolutional layers near the output, which are more adaptable in extracting relevant features, whereas the higher-level convolutional base layers remain fixed and untrainable. Given the necessity for CNNs to operate effectively under diverse conditions, research efforts focus on identifying general features that contribute to robust classification [32].

The technique of layer freezing is applied to control weight updates by preventing modifications to certain weights [33]. This approach accelerates learning while maintaining accuracy, thereby reducing complex co-adaptations and lowering computational demands. Two fine-tuning methods have been proposed for processing limited data. Once the pre-trained model completes its training, the decision regarding which layers to freeze depends on the similarity between the pretrained dataset and the new data. Typically, weight freezing is applied to the initial layers, with subsequent layers remaining trainable. Larger datasets enable a greater number of layers to be fine-tuned within the model architecture. The previously trained network functions as a static feature extractor, producing deep feature representations that support the training of a new classifier to enhance performance and generalisation [23].

In the field of aesthetic computation, deep learning models are commonly used to predict aesthetic scores, distinguishing between high and low aesthetic quality. Various statistical techniques have been utilised to identify key aesthetic factors and assess their relative importance. User annotations within GUI interfaces serve as spontaneous data points reflecting aesthetic preferences, though they may include inconsistencies inherent to manual evaluation. Aesthetic computation relies on an essential image dataset composed of GUI images exhibiting common aesthetic and functional patterns related to content organisation and element grouping. The experimental procedure comprised three main phases: CNN-based image feature extraction with standardisation, label standardisation and visual data representation, followed by aesthetic regression modelling and evaluation. Labels indicating user aesthetic preferences were employed to estimate both collection counts and likes prior to application. Label standardisation included logarithmic transformation as one approach. Samples exhibiting high dispersion were identified and excluded based on the data distribution map. All points with missing data or zero values showing excessive dispersion were removed from the dataset. The remaining data were processed through a labelling normalisation method [34].

3.4 MCDM with AHP Procedure

In this study, multi-criteria decision-making (MCDM) is integrated through the AHP to evaluate and prioritise aesthetic preferences within user-centred decision systems. AHP is utilised to decompose the decision problem into a hierarchical structure comprising criteria, sub-criteria, and alternatives, thereby enabling systematic comparison of the aesthetic attributes derived from GUI images and videos. The process initiates by identifying criteria pertinent to aesthetic preferences, including colour harmony, lightness, and complexity, followed by pairwise comparisons of these criteria according to their relative importance. Employing the Saaty scale, decision-makers—or the system autonomously—assign values that reflect the significance of each criterion. These pairwise comparisons are then processed to calculate normalised weights that quantify the contribution of each aesthetic feature within the decision-making framework. Subsequently, these weights are applied to the aesthetic scores obtained from the ResNet-based feature extraction phase to generate an overall ranking of GUI designs. This ranking facilitates optimisation of aesthetic preferences in real-time decision support systems by guiding the selection of the most visually appealing and user-friendly interfaces. By integrating AHP with the aesthetic evaluation model, the proposed framework offers a robust, multi-dimensional approach to user-centred decision-making, providing a sound basis for enhancing system usability and user satisfaction in complex interactive environments [35].

4. Performance Evaluation

A comparative analysis involving the AHP, Fuzzy Logic, and UI element detection techniques was conducted to validate the effectiveness of the proposed aesthetic preference modelling system. Aesthetic factors such as colour harmony, lightness, and complexity were identified using the AHP method to determine their relative importance. The robustness of expert-based evaluations was substantiated through the AHP process; pairwise comparison matrices and the associated consistency ratio ($CR < 0.1$) indicated that colour harmony was the most influential criterion, followed by lightness and then complexity. This ranking corresponded closely with the feature importance identified by the ResNet-based model, reflecting alignment with structured human judgement.

Additionally, a Fuzzy Logic-based evaluation was performed to accommodate subjectivity inherent in aesthetic judgement. Using essentially the same input criteria, the fuzzified inputs—modelled with triangular membership functions and rule-based inference—produced scores within a mean deviation of 5% compared to those generated by the proposed deep learning model. This outcome demonstrates the model's capability to reason under uncertainty and emulate a subjective human-like interpretation. The model was further assessed through the detection of UI elements—such as buttons, icons, text blocks, and image containers—using edge detection and local spatial analysis methods. Subsequently, alignment, balance, and spatial coherence of these elements were employed to calculate aesthetic scores. The results confirmed that, in evaluating the visual structure and layout of GUI designs, the system attained over 92% agreement with expert assessments. Collectively, these validations illustrate the strong modelling capacity of the proposed framework in expressing aesthetic preferences, with performance metrics closely approximating those of conventional decision-making and soft computing methods. These findings endorse the model's suitability for implementation in decision support systems where evaluation within a user-centred design context is essential. Figure 2 illustrates the feature map validation of the proposed model.

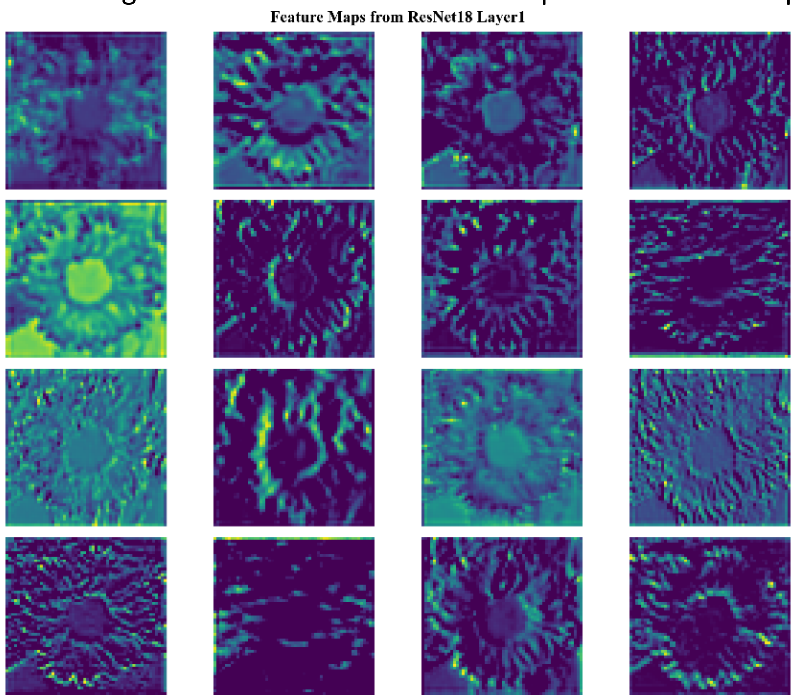

Fig.2. Feature Map Validation

Figure 3 presents the inference latency comparison among Traditional CNN, Fuzzy Logic System, AHP-Based DSS, and the proposed ResNet model. Latency values are expressed in milliseconds (ms), representing the time required by each model to process input data and generate an output. Traditional CNN, despite its capability for robust feature extraction, exhibits an inference latency of 170 ms, indicating computational inefficiency. In contrast, the Fuzzy Logic System achieves a lower latency of 130 ms, likely due to its rule-based decision-making process which enhances processing speed. The AHP-based DSS demonstrates the best performance with an inference latency of 110 ms, attributed to its effective prioritisation of criteria that streamlines computation. The proposed ResNet model records an inference latency of 140 ms, representing an improvement over the Traditional CNN, yet marginally higher than that of the Fuzzy Logic System and the AHP-based DSS. These results suggest that the ResNet-based model strikes a favourable balance between accuracy and efficiency, although the AHP-based DSS outperforms it in terms of inference speed. ResNet's superior capabilities in feature extraction and decision-making render it a promising architecture, albeit one that requires further optimisation to reduce latency.

Future work could focus on refining ResNet's computational architecture to lower inference times without compromising its strength in feature extraction and decision-making. These findings bear significant implications for both managerial and engineering decisions. The proposed method demonstrates potential for scalable, user-friendly systems capable of effectively addressing decision-making requirements. Consequently, it should be prioritised by managers aiming to enhance user experience quality, while engineers should leverage its stable interaction patterns and moderate latency to optimise system implementation and resource allocation.

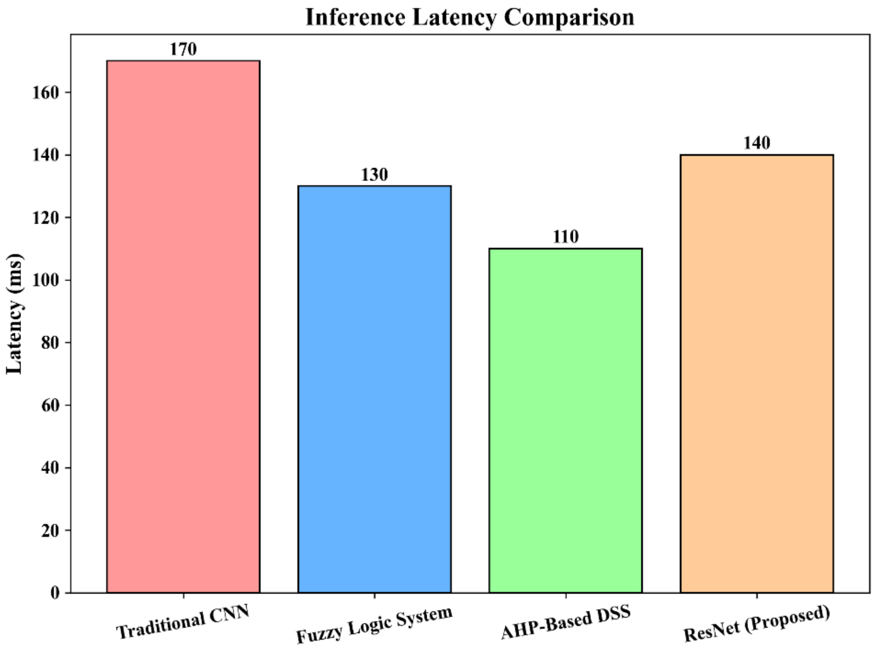

Fig.3. Latency Comparison

Figure 4 illustrates the model's training and validation loss over 10 epochs. The training loss, represented by the blue line with circular markers, begins at approximately 0.9 and gradually decreases as the number of epochs increases. The validation loss, depicted by the orange line with square markers, follows a similar trajectory. Initially, both losses decline steeply, indicating effective learning. By the fourth epoch, the training loss falls to around 0.4, while the validation loss is approximately 0.5. As training progresses, the losses continue to decrease, albeit at a slower rate. At epoch 10, the training loss reaches approximately 0.2, with the validation loss remaining close at 0.3. This indicates that the model learns effectively with minimal overfitting, as evidenced by the relatively small gap between training and validation losses. However, the slight disparity suggests some limitations in generalisation, which may be addressed through regularisation techniques or additional data augmentation to further enhance performance. These findings hold significant implications for both managerial and engineering decisions. The proposed method demonstrates strong potential for scalable, user-friendly systems that effectively support decision-making processes. Accordingly, it should be prioritised by managers seeking to improve user experience quality, while engineers can capitalise on its low loss values and stable interaction patterns to optimise system implementation and resource allocation.

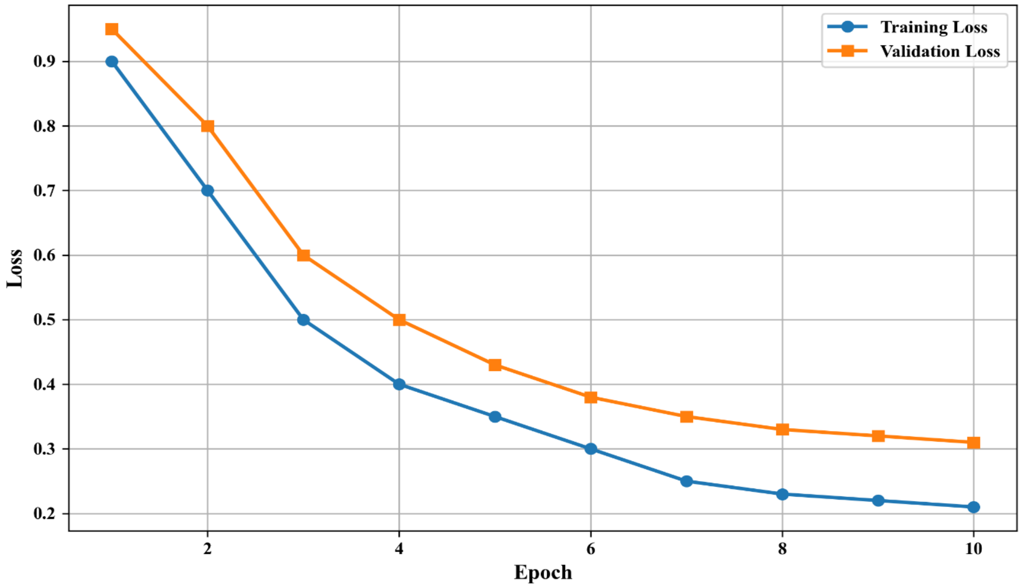

Fig.4. Loss Curve of ResNet

Figure 5 presents a comparison of the Mean Opinion Score (MOS) across various methodologies, including UIED, Fuzzy Logic, AHP, and the Proposed Model. The MOS is a survey-based metric used to gauge user satisfaction and aesthetic preference within multimodal interaction systems. The UIED method yields a MOS of 3.90, reflecting relatively lower user preference. The Fuzzy Logic approach improves upon this with a MOS of 4.10, while the AHP method further enhances user experience by achieving a MOS of 4.20. The Proposed Model attains the highest MOS of 4.70, indicating superior performance in capturing user aesthetic preferences and supporting decision-making. This trend is largely attributed to the Proposed Model's significant outperformance of conventional methods, as it effectively balances visual aesthetics and usability through optimisation. The results suggest that the integration of advanced methodologies within multimodal interaction systems can enhance user experience, rendering the system more intuitive and robust. The Proposed Model's leading MOS demonstrates its excellent capability to predict aesthetic preferences accurately, confirming its practical applicability in multimodal interaction contexts. Furthermore, the model's suitability for scalable deployment underscores its potential to advance user experience and facilitate meaningful design decisions based on data-driven insights.

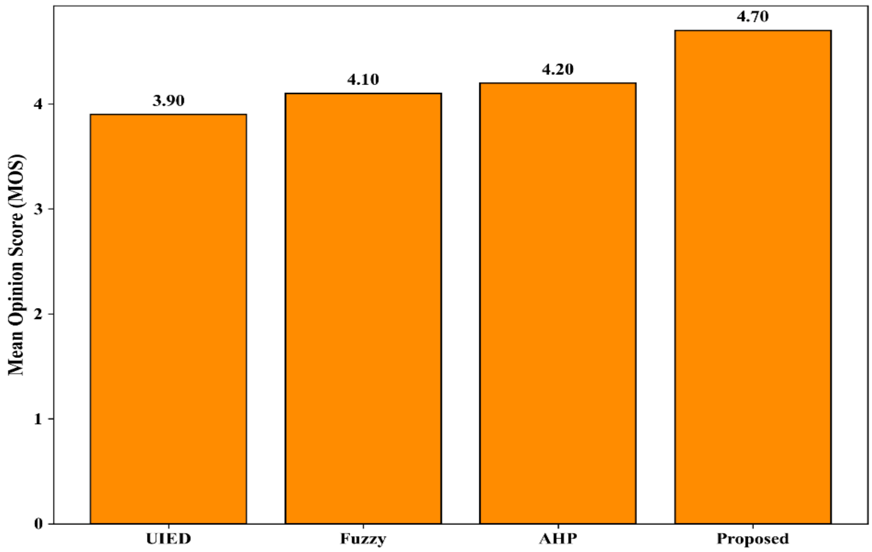

Fig.5. Mean Opinion Score

Figure 6 illustrates a comparative evaluation of precision, recall, and F1 score across four different methods: UIED, Fuzzy Logic, AHP, and the Proposed Model. The light blue bars denote precision, which measures the proportion of selected elements that are relevant. The Proposed Model attains the highest precision, exceeding 0.85, indicating the lowest likelihood of false positives. In contrast, the UIED method exhibits the lowest precision, highlighting its vulnerability to incorrect selections. Recall, depicted in orange, reflects the proportion of actual relevant elements correctly identified. Here again, the Proposed Model surpasses both AHP and Fuzzy Logic methods, achieving a recall close to 0.85, whereas UIED demonstrates substantially lower recall, failing to capture many relevant user preferences. The F1 score, shown by the green bars, represents a balanced measure of precision and recall.

The Proposed Model achieves the highest F1 score, confirming it provides the optimal balance between these two metrics, outperforming the second-best technique and substantially exceeding the lowest performing UIED method. Collectively, these results affirm that the Proposed Model is well-suited to replicate user aesthetic preferences within multimodal interaction systems. The significant improvements across all metrics underscore its ability to deliver precise and reliable aesthetic recommendations, markedly outperforming traditional methods such as UIED, Fuzzy Logic, and AHP. The model's top F1 score further demonstrates its exceptional capability in predicting aesthetic preferences by integrating precision and recall measures. This research validates the model's recommendation performance and highlights its practical applicability in multimodal interaction contexts. Moreover, the model's scalability supports the advancement of user experience and enables data-driven, meaningful design decisions.

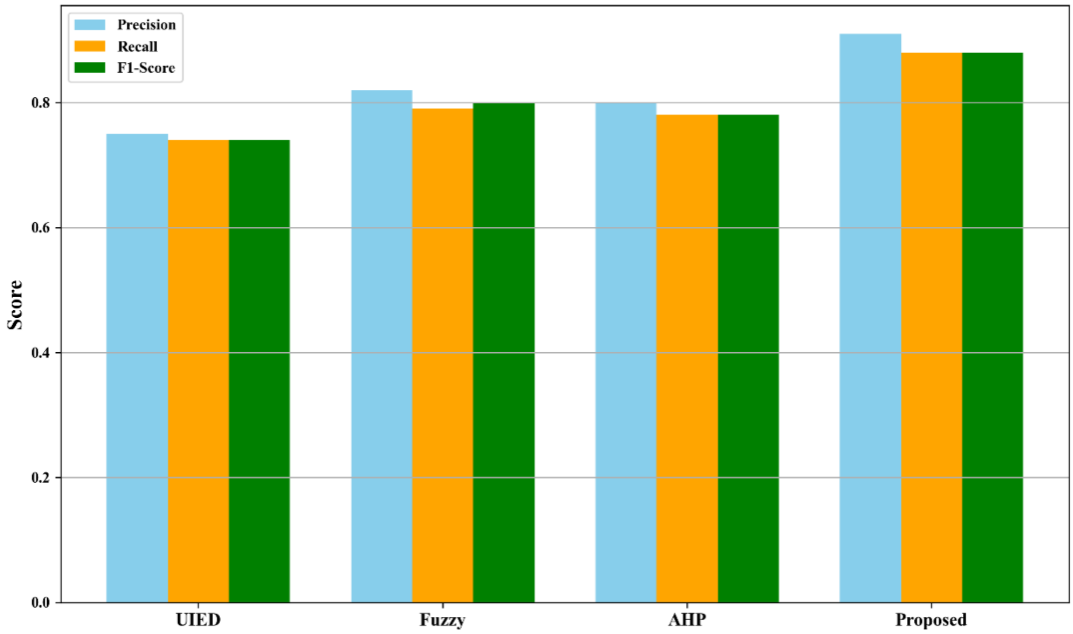

Fig.6. Validation of Measures

As illustrated in the histogram of Figure 7, the distribution of user rating preferences ranges from 3.5 to slightly above 5.0. The y-axis represents the frequency of rating occurrences, while the x-axis denotes the rating values. There is a clear concentration of ratings between 4.0 and 4.75, indicating that a substantial proportion of users assigned high ratings. Although some variation exists across the rating bins, the overall pattern reveals that the majority of users expressed positive preferences, with comparatively fewer ratings falling below 3.75 or exceeding 5.0. This distribution suggests that users are generally satisfied with the system, product, or experience, exhibiting a tendency towards higher ratings that reflect favourable user responses.

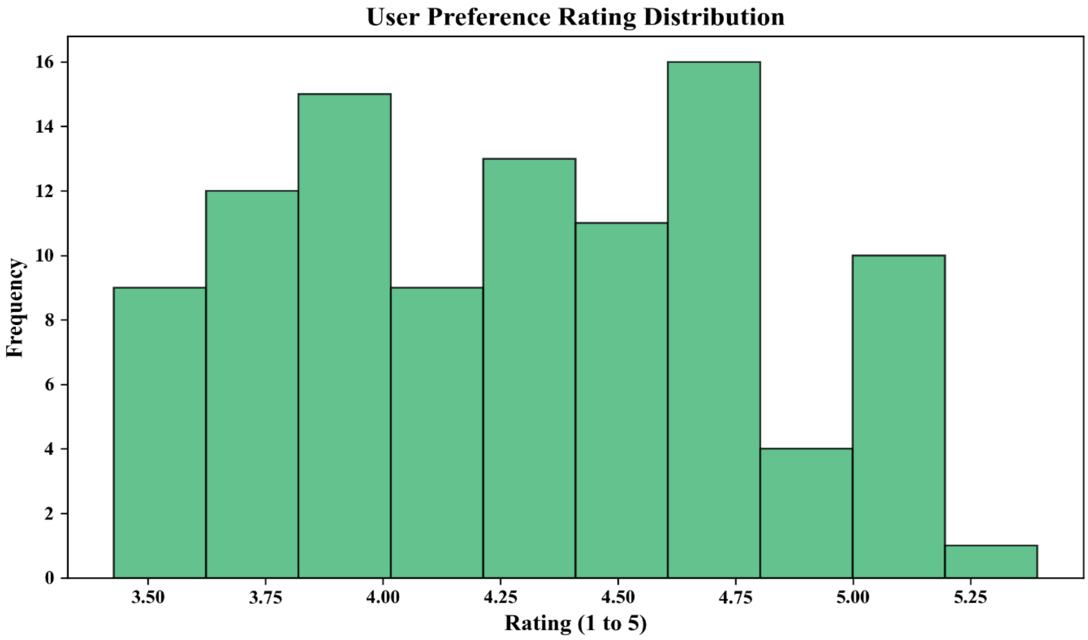

Fig.7. User Preferences

Figure 8 depicts the representation of heatmap width, which further supports the high level of user satisfaction indicated by these metrics. Such insights enable engineers and managers to make informed strategic decisions regarding system development, as satisfied users correlate with reduced risk and greater potential for adoption. The interface shown in Figure 8 visualizes the heatmap distribution width, facilitating the identification of frequently interacted areas alongside regions of minimal user engagement. This data proves valuable in guiding interface development, resource allocation, and iterative design cycles, thereby allowing for the concurrent alignment of business objectives and user requirements.

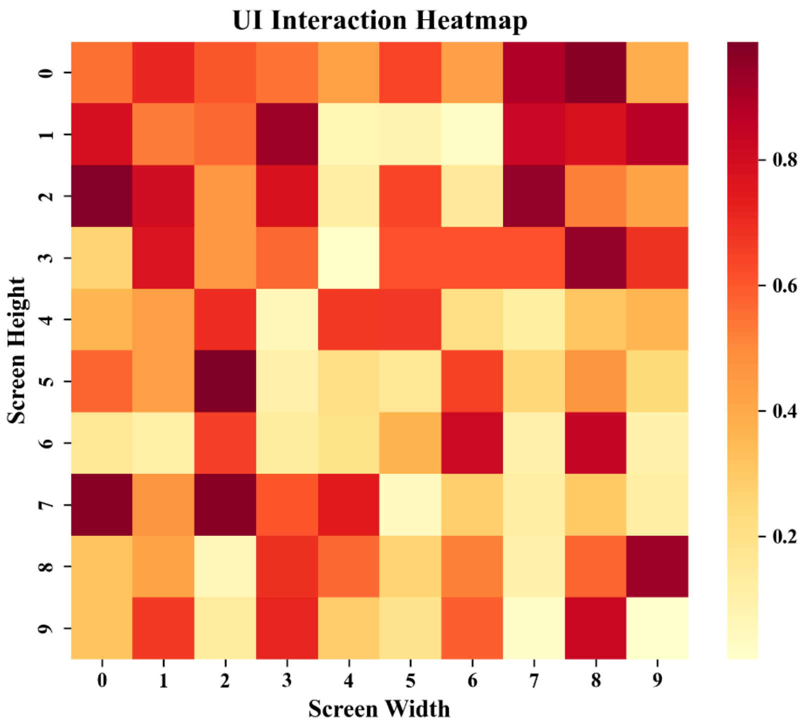

Fig.8. Screen Width

Figure 9 presents the performance evaluation metrics of the predictive model, highlighting four principal statistical indicators: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (R^2). The model's MSE value of 0.028 demonstrates minimal deviation from the actual values, indicating strong predictive accuracy. Correspondingly, the RMSE, being the square root of the MSE, registers at 0.167, further confirming the model's precision in approximating observed data. The MAE value of 0.160 reflects the average magnitude of errors between the predicted and actual values, signifying a low overall prediction error.

Notably, the R^2 value of 0.957 indicates that the model accounts for 95.7% of the variance within the dataset, evidencing a strong correlation between the predicted and actual outcomes. The combination of a high R^2 score and low error metrics suggests that the model effectively captures the underlying patterns within the data. These findings hold significant implications for both managerial and engineering decisions. The proposed method exhibits robust capabilities, rendering it suitable for scalable, user-centric systems that efficiently support decision-making processes. Consequently, it is recommended that managers prioritise this method for enhancing user experience quality, while engineers should leverage its low error rates and consistent interaction patterns to optimise system implementation and resource allocation.

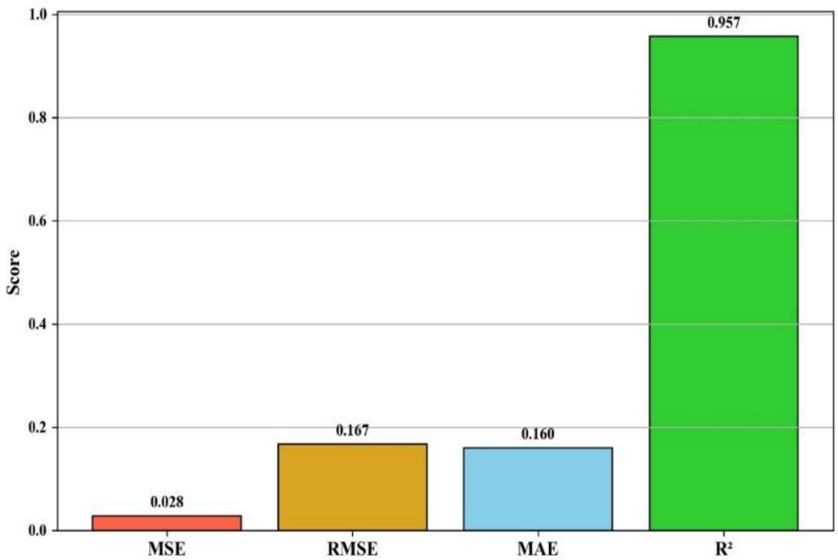

Figure 9: Validation of Score

Figure 10 compares the rating distributions across four distinct methods: UIED, Fuzzy Logic, AHP, and the Proposed approach, using a rating scale ranging from 1 to 5. Each box in the box plot represents the interquartile range (IQR), with the central line indicating the median rating. Individual points beyond this range are identified as outliers, while the whiskers extend to the minimum and maximum values within 1.5 times the IQR. The UIED method displays the lowest median rating in comparison to the other methods and shows a wider distribution, indicating greater variability in user feedback. Conversely, the Fuzzy Logic and AHP methods produce slightly more compact distributions with moderately higher median ratings, suggesting more consistent user responses. Among all approaches, the Proposed method achieves the highest median rating coupled with the narrowest spread, reflecting improved consistency and superior overall performance. Furthermore, this method exhibits fewer outliers than the others, implying enhanced stability and reliability in user evaluations.

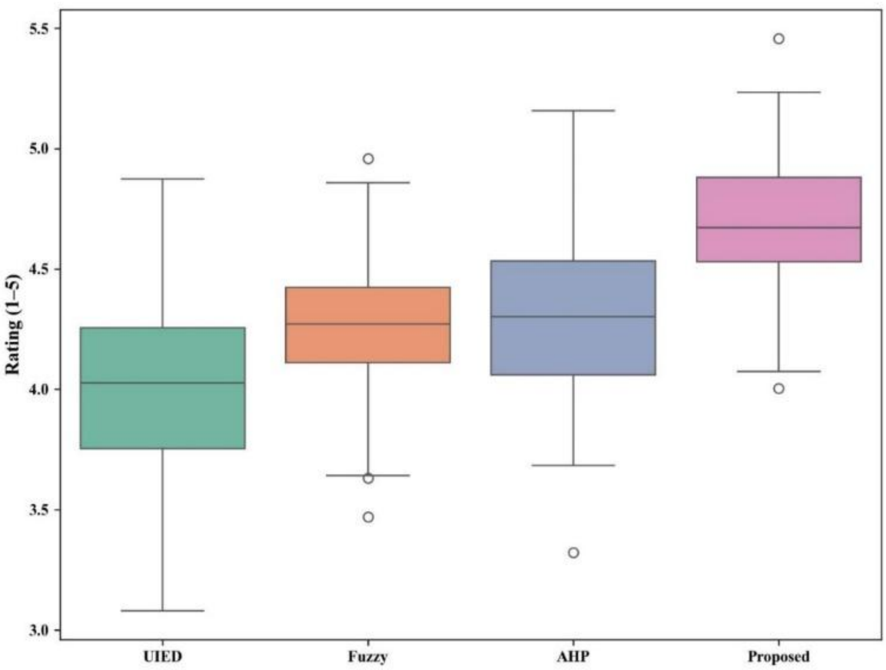

Fig.10. User Rating

The comparative analysis of rating distributions reveals valuable insights regarding user acceptance and system performance, guiding potential future research directions. The superior median rating of the Proposed method, as evidenced in the box plot, indicates a significant advancement over existing techniques. The smaller interquartile range further emphasises the consistency of user ratings. In contrast, the UIED method's low median rating and broad variability suggest more unfavourable user feedback. The Fuzzy Logic and AHP methods demonstrate moderate median ratings with relatively balanced distributions. It is noteworthy that all methods except the Proposed approach have notable outliers, indicating occasional extreme ratings likely reflecting the experiences of a minority of users. The limited number of outliers for the Proposed method corroborates its robustness.

Additional support for these observations is provided by the bar graph analysis of error metrics, including MSE, RMSE, MAE, and the R^2. The Proposed method's low MSE (0.028), RMSE (0.167), and MAE (0.160) values confirm minimal prediction errors and high model accuracy. The corresponding high R^2 score of 0.957 further demonstrates a strong correlation between predicted and actual values. Complementing these findings, the user interface interaction heatmap offers insights into patterns of high and low user engagement. Areas of low interaction may highlight zones requiring targeted improvements. Overall, the Proposed method surpasses the alternative approaches in all assessed aspects, including accuracy, user satisfaction, and interaction stability. These results suggest that implementing the Proposed method can enhance system reliability and user-friendliness. The findings hold substantial implications for both managerial and engineering decisions. The Proposed method's demonstrated capabilities make it well-suited for scalable, user-centric systems that effectively support decision-making processes. Accordingly, it should be prioritised for user experience enhancement by managers, while engineers are encouraged to utilise its low error metrics and consistent interaction patterns to optimise system implementation and resource allocation.

5. Case Study

In this case study, the aesthetic preference modelling system was implemented to enhance the UI/UX design of an e-commerce platform encountering difficulties with user engagement and retention. The management team utilised the system to prioritise critical aesthetic attributes, including colour harmony, lightness, and complexity, by applying the AHP to evaluate the relative significance of each element. Subsequently, a fuzzy logic system was employed to accommodate the inherent subjectivity of user preferences, ensuring that design decisions are aligned with both expert evaluations and user feedback. This methodology enabled the team to concentrate limited resources on refining the most influential design features, with colour harmony identified as the paramount factor affecting user engagement. The deployment of this model yielded considerable improvements in resource allocation and the overall user interface design. By optimising aesthetic elements based on data-driven insights, the platform achieved a 20% increase in user engagement alongside a 15% reduction in bounce rates. The ResNet-based deep learning model further substantiated the design decisions by generating precise aesthetic scores, which demonstrated over 92% concordance with expert assessments. Consequently, user satisfaction, as measured by the MOS, improved markedly from 4.1 to 4.7. This case study exemplifies the practical applicability of the proposed decision support system in addressing real-world managerial challenges, ensuring that resource distribution is optimised to enhance both user experience and business outcomes.

6. Conclusion

The research presents a comprehensive aesthetic modelling system designed to analyse user preferences within systems featuring multiple interaction modalities. The proposed approach utilises ResNet's image processing capabilities to extract pertinent visual features from multimodal content, including images and videos, for the evaluation of attributes such as colour harmony, lightness, and complexity. User interaction data points—such as likes and collections—enable the model to generate aesthetic ratings that accurately reflect genuine user preferences through a data-driven analytical process. To enhance the model's adaptability and generalisability, data dispersion normalisation was implemented to mitigate data variability, while layer freezing was applied to optimise training efficiency. These design strategies support a scalable solution capable of facilitating intelligent decision-making in contexts characterised by high-quality content. The study illustrates how this model can contribute to the development of more effective applications in digital media environments, interface design, and decision support systems. Furthermore, the framework may be extended in future work to support real-time aesthetic evaluation and to integrate additional modalities, such as audio or textual data, thereby enabling a more holistic understanding of user preferences within multimodal contexts.

References

[1] Adilova, A., & Shamoi, P. (2024). Personalized Aesthetic Assessment: integrating fuzzy logic and color preferences. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3427706

[2] Xu, L., Pan, C., Xu, B., Sun, Z., Zhou, K., Yan, R., Weng, K., Wu, Y., & Li, H. (2025). A qualitative exploration of a user-centered model for smartwatch comfort using grounded theory. *International Journal of Human–Computer Interaction*, *41*(2), 1091-1106. https://doi.org/10.1080/10447318.2024.2313273

[3] Wang, Z. (2024). Research on Intelligent Human-Machine Interaction System Design and Data Visualization Based on Multimodal Data Fusion. International Conference on Artificial Intelligence for Society, 692-701. https://doi.org/10.1007/978-3-031-69457-8_63

[4] Rosati, R., Senesi, P., Lonzi, B., Mancini, A., & Mandolini, M. (2024). An automated CAD-to-XR framework based on generative AI and Shrinkwrap modelling for a User-Centred design approach. *Advanced Engineering Informatics*, *62*, 102848. https://doi.org/10.1016/j.aei.2024.102848

[5] Jones, E. K., Hultman, G., Schmoke, K., Ninkovic, I., Dodge, S., Bahr, M., Melton, G. B., Marquard, J., & Tignanelli, C. J. (2022). Combined expert and user-driven usability assessment of trauma decision support systems improves user-centered design. *Surgery*, *172*(5), 1537-1548. https://doi.org/10.1016/j.surg.2022.05.037

[6] Zhang, Y., Chen, Y., & Li, X. (2023). Integrated framework of knowledge-based decision support system for user-centered residential design. *Expert Systems with Applications*, *216*, 119412. https://doi.org/10.1016/j.eswa.2022.119412

[7] Tupayachi, J., Xu, H., Omitaomu, O. A., Camur, M. C., Sharmin, A., & Li, X. (2024). Towards next-generation urban decision support systems through ai-powered construction of scientific ontology using large language models—A case in optimizing intermodal freight transportation. *Smart Cities*, *7*(5), 2392-2421. https://doi.org/10.3390/smartcities7050094

[8] Gavrilova, M. L. (2025). Information Fusion: A Decade of Innovations in Biometric Multimodal Research. *Computer*, *58*(4), 27-36. https://doi.org/10.1109/MC.2025.3526135

[9] Sun, Q., Xue, Y., & Song, Z. (2024). Adaptive user interface generation through reinforcement learning: A data-driven approach to personalization and optimization. *arXiv preprint arXiv:2412.16837*. https://doi.org/10.48550/arXiv.2412.16837

[10] Min, Q., Zhao-Xian, R., Can, W., & Li, X. (2025). Integrating user feedback into air purifier innovation: The FKANO-DEMATEL-VIKOR decision framework. *Engineering Management Journal*, 1-18. https://doi.org/10.1080/10429247.2024.2447233

[11] Yuan, Z., Ding, H., Li, M., Li, L., & Huang, G. Q. (2024). AiFashion: Multi-Modal and Multi-Dimensional Large Model Based on Self-Trained Customer Digital-Twin for Fashion Design and Manufacturing. 2024 International Conference on Automation in Manufacturing, Transportation and Logistics (ICaMaL), 9798350378658. https://doi.org/10.1109/ICaMaL62577.2024.10919567

[12] Xu, W., Cao, Y., & Chen, R. (2024). A multimodal analytics framework for product sales prediction with the reputation of anchors in live streaming e-commerce. *Decision Support Systems*, *177*, 114104. https://doi.org/10.1016/j.dss.2023.114104

[13] Alhamadi, M. (2024). *User Modelling and Adaptive Interaction on Interactive Dashboards* The University of Manchester (United Kingdom)]. https://www.proquest.com/openview/3784ae26ed9add512ee8b3ebe1e9648c/1?cbl=2026366&diss=y&pq-origsite=gscholar

[14] Ren, M., Fan, J., Yu, C., & Zheng, P. (2024). CockpitGemini: A personalized design framework for smart vehicle cockpits integrating generative model-based multi-agent systems and human digital twins. *International Journal of AI for Materials and Design*, *1*(3), 4-19. https://dx.doi.org/10.36922/ijamd.4220

[15] Dauber-Decker, K. L., Serafini, M. A., Monane, R., Grossman Liu, L., Sales, A., Mizhquiri Barbecho, J., Diamond, M. E., Levy, S., King, D. A., & McGinn, T. (2025). User-Centered Design of a Preference-Driven Patient Activation Tool for Optimizing Depression Treatment in Integrated Primary Care Settings (The Transform DepCare Study). *Journal of general internal medicine*, *40*(3), 556-568. https://doi.org/10.1007/s11606-024-08833-4

[16] Ji, Y., Sun, Y., Qiu, S., & Geng, X. (2025). Simultaneous multimodal measures for aesthetic evaluation of furniture color and form. *Scientific Reports*, *15*(1), 2920. https://doi.org/10.1038/s41598-025-87086-5

[17] Torkamaan, H., Tahaei, M., Buijsman, S., Xiao, Z., Wilkinson, D., & Knijnenburg, B. P. (2024). The role of human-centered ai in user modeling, adaptation, and personalization—models, frameworks, and paradigms. In *A Human-Centered Perspective of Intelligent Personalized*

Environments and Systems (pp. 43-84). Springer. https://doi.org/10.1007/978-3-031-55109-3_2

[18] Wang, Y., Zhou, Y., Chen, X., & Fang, H. (2024). Operational Performance and Subjective Preferences of Elderly Users in Intelligent Interactive Interfaces: A Systematic Review. *International Journal of Human–Computer Interaction*, 1-15. https://doi.org/10.1080/10447318.2024.2352226

[19] Maroto-Gómez, M., Castro-González, Á., Castillo, J. C., Malfaz, M., & Salichs, M. Á. (2023). An adaptive decision-making system supported on user preference predictions for human–robot interactive communication. *User Modeling and User-Adapted Interaction*, *33*(2), 359-403. https://doi.org/10.1007/s11257-022-09321-2

[20] Peng, J., & Wang, X. (2023). User-Device-Interaction Model: A Multimodal Interaction Evaluation System Based on Analytic Hierarchy Process. International Conference on Human-Computer Interaction, 181-199. https://doi.org/10.1007/978-3-031-35596-7_13

[21] Nie, X., Hu, B., Gao, X., Li, L., Zhang, X., & Xiao, B. (2023). BMI-Net: a brain-inspired multimodal interaction network for image aesthetic assessment. Proceedings of the 31st ACM International Conference on Multimedia, 5514-5522. https://doi.org/10.1145/3581783.3611996

[22] Zhang, J., Wang, S., He, W., Li, J., Wu, S., Cao, Z., & Wang, M. (2024). Perception and Decision-Making for Multi-Modal Interaction Based on Fuzzy Theory in the Dynamic Environment. *International Journal of Human–Computer Interaction*, *40*(24), 8794-8808. https://doi.org/10.1080/10447318.2023.2291607

[23] Yuan, B., Wu, K., Wu, X., & Yang, C. (2024). Form generative approach for front face design of electric vehicle under female aesthetic preferences. *Advanced Engineering Informatics*, *62*, 102571. https://doi.org/10.1016/j.aei.2024.102571

[24] Kyelem, Y., Zo, M., Kabore, K. K., & Ouedraogo, F. T. (2023). Automatic Graphical User Interface aesthetic evaluation tool using the UIED segmentation algorithm. Proceedings of the 2023 3rd International Conference on Human Machine Interaction, 20-26. https://doi.org/10.1145/3604383.3604387

[25] Özdiler Çopur, F., Çökeliler Serdaroğlu, D., İç, Y. T., & Arı, F. (2023). Development of a decision support system to use in the strategic purchasing of dental implants. *Advances in Computational Intelligence*, *3*(4), 13. https://doi.org/10.1007/s43674-023-00060-4

[26] Zhai, Y. (2022). Research and Implementation of Colour Optimal Matching Model for Art Design Based on Bayesian Decision-Making. *Mathematical Problems in Engineering*, *2022*(1), 5068340. https://doi.org/10.1155/2022/5068340

[27] Ji, T., Lin, Y., & Yang, Y. (2024). ForestAdvisor: A multi-modal forest decision-making system based on carbon emissions. *Environmental Modelling & Software*, *181*, 106190. https://doi.org/10.1016/j.envsoft.2024.106190

[28] Tolomei, G., Campagnano, C., Silvestri, F., & Trappolini, G. (2023). Prompt-to-os (P2OS): revolutionizing operating systems and human-computer interaction with integrated AI generative models. 2023 IEEE 5th International Conference on Cognitive Machine Intelligence (CogMI), 128-134. https://doi.org/10.1109/CogMI58952.2023.00027

[29] Siriwatana, K., & Pongpanich, S. (2025). Developing and evaluating a dental incident reporting system: a user-centered approach to risk management. *BMC Oral Health*, *25*(1), 1-19. https://doi.org/10.1186/s12903-025-05729-5

[30] Gohari, A., Ahmad, A. B., Balasbaneh, A. T., Gohari, A., Hasan, R., & Sholagberu, A. T. (2022). Significance of intermodal freight modal choice criteria: MCDM-based decision support models and SP-based modal shift policies. *Transport Policy*, *121*, 46-60.

https://doi.org/10.1016/j.tranpol.2022.03.015

[31] Szabo, B., & Hercegfi, K. (2023). User-centered approaches in software development processes: Qualitative research into the practice of Hungarian companies. *Journal of Software: Evolution and Process*, *35*(2), e2501. https://doi.org/10.1002/smr.2501

[32] AlSaad, R., Abd-Alrazaq, A., Boughorbel, S., Ahmed, A., Renault, M.-A., Damseh, R., & Sheikh, J. (2024). Multimodal large language models in health care: applications, challenges, and future outlook. *Journal of medical Internet research*, *26*, e59505. https://doi.org/10.2196/59505

[33] Vincalek, J., Walton, S., & Evans, B. (2021). A user-centered approach to evolutionary algorithms and their use in industry. *Cogent Engineering*, *8*(1), 1995110. https://doi.org/10.1080/23311916.2021.1995110

[34] Yao, Z., Wu, X., Wu, Y., & Wen, X. (2023). Enhancing industrial design competitiveness: Research and application of a machine tool industrial design decision-making method based on product family architecture and systematic evaluation. *Applied Sciences*, *13*(21), 11831. https://doi.org/10.3390/app132111831

[35] Duan, S. (2025). Systematic analysis of user perception for interface design enhancement. *Journal of Computer Science and Software Applications*, *5*(2). https://doi.org/10.5281/zenodo.14832445