

Aesthetics Prediction of Websites using Federated Learning

Ranjan Maity

`r.maity@cit.ac.in`

CITK: Central Institute of Technology Kokrajhar Deemed to be University <https://orcid.org/0000-0002-6865-1067>

Jahnvi Roy

Central Institute of Technology Kokrajhar

Rahul Son Boro

Central Institute of Technology Kokrajhar

Deepali Borgayary

Central Institute of Technology Kokrajhar

Research Article

Keywords: Federated Learning, User Interface, UI Aesthetics, Deeplearning, EfficientNet-B0, Personalised Aesthetic Prediction.

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11396505/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Aesthetics Prediction of Websites using Federated Learning

Jahnavi Roy¹, Rahul Son Boro¹, Deepali Borgayary¹, and Ranjan Maity¹

Department of Computer Science and Engineering
Central Institute of Technology Kokrajhar, Kokrajhar, India
email: r.maity@cit.ac.in

Abstract. Interface aesthetics is considered subjective in nature. Most existing approaches to aesthetic prediction generally rely on centralized data collection and use average ratings, often ignoring individual user choices. To address this, we propose a Federated Learning-based model. To construct the base model, we considered EfficientNet-B0, ResNet-50, and Vision Transformer (ViT-B/16). To test these three models, we developed a dataset of 100 website interfaces (50 mobile and 50 desktop/laptop) and had 100 users rate them using the Latin Square Method. It was observed that EfficientNet-B0 outperformed the others. Consequently, we considered EfficientNet-B0 as the base model. In the empirical study, we found that users can be grouped into five types based on their gender and occupation. Accordingly, five clients were considered for our Federated Learning algorithm. The Federated Average (**FedAvg**) algorithm achieved a promising MSE of only 0.0534.

Keywords: Federated Learning, User Interface, UI Aesthetics, Deep-learning, EfficientNet-B0, Personalised Aesthetic Prediction.

1 Introduction

The visual aesthetics in user interfaces (UI) play an important role in determining users' satisfaction, engagement, and how they perceive a product overall [1]. Users decide whether they like an interface in the first 50 milliseconds [2]. These initial impressions can be a major factor in their decision to keep using the app or throw it away. Since mobile and web applications are becoming more and more widespread, the creators and builders need to come up with UIs that are not only effective but also nice to look at.

Nevertheless, assessing aesthetics is a profoundly subjective activity because of human nature [3]. People may have different choices due to their cultural backgrounds, ages, and genders. Nowadays, most of the methods used to measure the aesthetic quality of a user interface (UI) are human-based, such as conducting surveys, hiring experts to review, or conducting heuristic evaluations [4]. These methods are useful; however, they are not only time-consuming but also very limited in their ability to be scaled. Additionally, they may even be influenced by the individual's bias, which consequently makes them unusable in cases where the number of designs is very large.

Machine learning [5–7] and deep learning [8] have the potential to become the tools for automating aesthetic evaluation. Convolutional Neural Networks (CNNs) [9] are known for their ability to capture high-level visual characteristics of images, and they have also been used successfully in image aesthetic assessment. However, the majority of the current models come up with only one average aesthetic score that indicates the preference of the general population, thereby disregarding the fact that aesthetic perception varies from one individual to another.

In this work, we address this gap by proposing a personalized aesthetic score prediction system for user interfaces using Federated Learning. Rather than learning a single global aesthetic model. The adoption of Federated Learning in this work is motivated not solely by privacy considerations, but by a set of deeper technical and architectural reasons rooted in aesthetic data itself. Human aesthetic perception is very personal and very different from one person to another. As a result, rating distributions among users are very non-IID (Non-Independent and Identically Distributed). Federated Learning is therefore an ideal approach for addressing such challenges. By enabling each client to learn and preserve local preferences while simultaneously contributing to the global model, the framework supports personalization at scale and reduces the bias introduced by globally averaged patterns. Moreover, it facilitates continuous adaptation to evolving user preferences without requiring centralized data collection or pooling.

Besides that, Federated Learning fits very well with the idea of running systems at the edge, and that is why the proposed method is suitable for on-device personalization in browsers, mobile apps, as well as design assistant tools. The framework also ensures that the raw user data is never transferred from the local device, which not only safeguards the user’s privacy but also avoids the centralization of sensitive rating data. In this work, we considered the web interfaces of the fifty websites. For each interface, two versions—mobile and desktop/laptop—are considered. The base model of esthetics was designed using EfficientNet-B0. Following this, the federated average algorithm was used. Experimental results show that the proposed approach may be a potential candidate for aesthetics computation.

2 Related Work

Over the years, researchers have tried to measure the subjective property of interface usability. It was observed that there is significant scope to improve the interface’s usability [10]. Usability is generally measured by five parameters: efficiency, learnability, memorability, error rate, and visual appearance (aesthetics). Existing works attempt to model aesthetics of an interface as image aesthetics [11, 12]. On the contrary, there are others where interface aesthetics is measured by the components’ (text [13], image [5], short animation [14]) aesthetics of an interface and their geometrical properties [15]. In the following, we list some relevant works related to our study.

The **AVA** dataset [16] is considered a comprehensive dataset for image aesthetic research. The dataset contains more than 2,50,000 images and their ratings on a 10-point Likert scale. The existing computational models of aesthetics can be classified broadly into two types - Machine Learning-based and Deep Learning-based approaches. A brief review of them is reported in the following.

Machine Learning-based approaches use handcrafted features to develop the computational model for aesthetics prediction. It was reported that an interface is generally composed of objects. Their geometrical positions - wire-frame geometry are responsible parameters for aesthetics prediction [6, 15, 17]. The geometrical parameters used here are - *Balance, Cohesion, Density, Equilibrium, Economy, Homogeneity, Proportion, Regularity, Rhythm, Sequence, Simplicity, and Symmetry*. Except for the wireframe-based approaches, researchers considered the contents of an interface as a key to measuring its aesthetics. Datta et al. [18] work on image aesthetics assessment can be considered as one of the early works in this direction. Following this, several works reported modeling interface aesthetics as image aesthetics [5]. The list of the features used in these studies is mainly *contrast, hue, saturation, value, smoothness, aspect ratio, unique color, sharpness*, and the *rule of thirds*.

Machine learning-based approaches are also used to model the aesthetics of the text components present in an interface [13]. Different handcrafted text features - *Font Size, Line Height, Chromatic Contrast, Luminance Contrast, Word Spacing, Letter Spacing* were used in this study. In all the existing works, different machine learning techniques - *non-linear regression* [13], Support Vector Machine classification [6], Support Vector Regression [1] were used. Unlike structured data (in the form of feature values), deep learning techniques can effectively figure out the implicit features from unstructured data - image, audio, and text [19]. Over the years, several deep learning models have been proposed to measure aesthetics. Most of these existing works were carried out to measure image aesthetics [20, 21]. A brief review of them is reported in [8]. Even though there has been a lot of progress in machine learning and deep learning, there are still scopes to address the following:- Research on the aesthetic evaluation of GUI is limited, particularly regarding differences between desktop and mobile interfaces. Assessing user interface (UI) aesthetics is inherently subjective, and there are no rigorously validated computational models that reliably evaluate aesthetics while accounting for individual differences among users. Most contemporary models of aesthetic prediction rely on ground-truth labels derived from mean ratings aggregated at the population level, thereby disregarding substantial inter-individual variability in aesthetic experience that arises from factors such as age, gender, and educational background.

3 Data Collection

To address the issues above, we note that there is no comprehensive dataset of webpage aesthetics available. Consequently, a dataset using one hundred webpages was developed.

3.1 Website Selections

A total of one hundred UI screenshots were collected, consisting of 50 desktop versions, and 50 mobile versions of the same interfaces. The interfaces were selected from ten different domains - Government official websites, Sports platforms, Job portals, News websites, OTT platforms, Courier services, Educational platforms, Travel websites, Environmental platforms, and Shopping applications to ensure diversity in visual design and usability characteristics. In Figure 1, a sample - desktop and mobile version of the *Sewa Setu* website¹ is shown. Screenshots taken of desktop applications were at a resolution of 1920 (horizontal) $\times$ 1080 (vertical) pixels (full-HD), and those for mobile devices were at 640 (horizontal) $\times$ 1136 (vertical) pixels (standard smartphone).

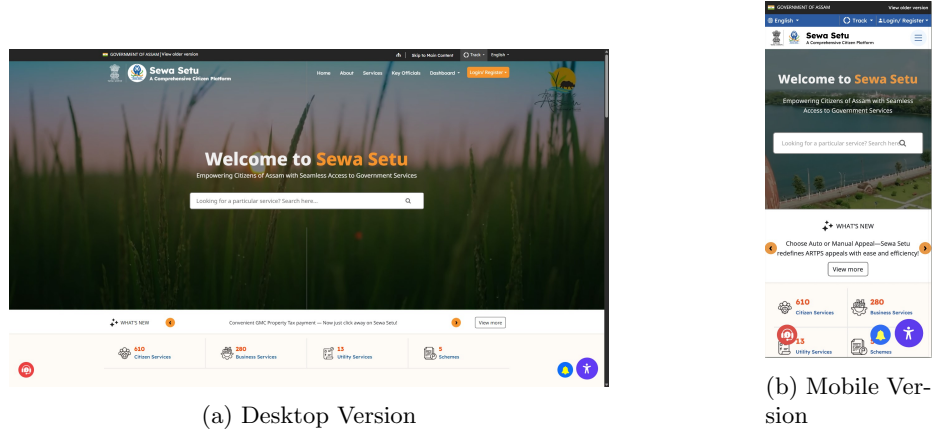


Fig. 1: Screenshot of Sewa Setu -(a) Desktop/Laptop (b) Mobile version.

3.2 Empirical Data Collection

In this study, one hundred participants were used to rate the one hundred images (**within group**) on a ten-point Likert scale. Subjects consisted of: 32 individuals aged between 6 - 17 years, 51 individuals aged between 18 to 30 years, and 17 individuals over the age of 31 years. The breakdown of occupation consisted of: working, non-working, and students. A sample table consisted of both males and females, where male = 0, female =1, and non-working users : 0, students: 1, working users: 2 is reported in Table ???. All the participants rated one hundred images using the *Latin Square* method. This was done to avoid any biases that may arise during the empirical study. Altogether one lakh data (one hundred interfaces rated by one hundred users) was collected. We were interested in

¹ <https://sewasetu.assam.gov.in/>

checking whether aesthetics scores vary from mobile interfaces to non-mobile interfaces. In the following, a statistical analysis of the empirical data is reported.

3.3 Paired t-test

A paired t-test is often used to check the statistical difference between the means of two related groups. In this work, the two groups are the desktop ratings and the mobile ratings of aesthetic. Accordingly, we formed the following null hypothesis -

$$\mu_{desktop_aesthetics} = \mu_{mobile_aesthetics} \quad (1)$$

$\mu_{desktop_aesthetics}, \mu_{mobile_aesthetics}$ denotes the mean of desktop and mobile aesthetics, respectively. The paired t-test was conducted using IBM SPSS. The following result was noted -

t-test Result
$\mu_{desktop_aesthetics}=6.039$ $\mu_{mobile_aesthetics}=6.084$ $P_{value} = 0.267, P_{value} > 0.05$

From the above-mentioned results, it may be noted that the observed $P_{value} = 0.267$ is greater than the 95% significance level (0.05). Hence, we can claim that there is no significant difference in aesthetics between the desktop and the mobile version. Using the collected data of one lakh, we planned to develop our proposed model of Federated Learning (FL). To develop an FL model, there is a need to develop a centralized base model. In the following, we discussed it.

4 Model Development

4.1 Base Model Development

We considered EfficientNet-B0 [22] as our base model. It is considered to be a fast model having fewer parameters and lower computational resources, while it maintains scalability and transfer learning capabilities. We used the pre-trained weights of ImageNet-1K, and replaced its final classifier layer with a single linear unit to adapt it for regression (as we were intersted to predict aestheics as a continuous quantity). The model was trained using the *AdamW optimiser* with a learning rate of 1×10^{-4} and a batch size of 16. A *ReduceLROnPlateau* scheduler was used to reduce the learning rate by a factor of 0.5 when validation performance stopped improving for 3 consecutive epochs. *Automatic Mixed Precision* (AMP) accelerated GPU computations and lowered memory usage. The training ran for up to fifty epochs with early stopping triggered when the validation loss failed to improve for 10 consecutive epochs. All losses are reported on the 0 to 1 normalised rating scale (personalised and min-max scaled). As shown in Figure 2, the training loss dropped sharply from approximately 0.098 at epoch 1 to around 0.064 by epoch 5, then gradually settled near 0.061 by epoch 20. The validation loss decreased quickly in the first few epochs and

thereafter remained stable around 0.054, closely tracking the training curve. The narrow gap between training and validation loss indicates good generalisation without significant overfitting. The mild, persistent divergence may reflect the lightweight design of EfficientNet-B0, which occasionally leaves a small amount of the aesthetic variance unexplained, though it still delivers strong and reliable predictions. This combination of smooth convergence, low memory footprint, and robust validation performance made EfficientNet-B0 the preferred backbone for the subsequent federated learning experiments. Following this, we proposed our FL-based architecture as reported below.

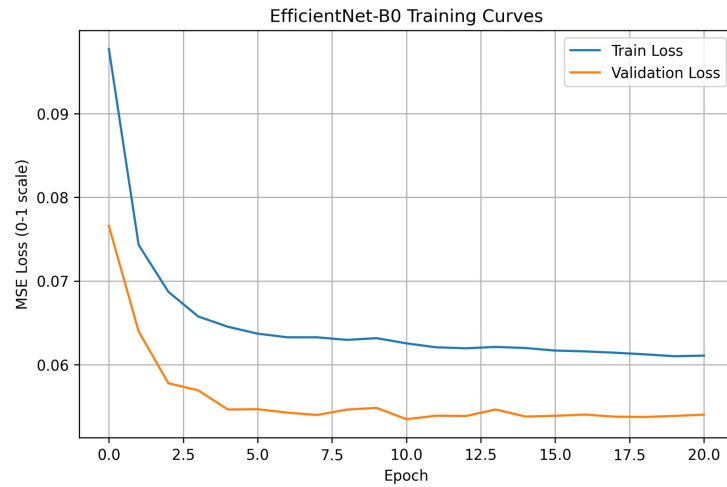


Fig. 2: Training and Validation Loss of EfficientNet

4.2 Federated Model Architecture

This section details the preprocessing pipelines, the neural network architecture, and the training setup that constitute the final model used in the federated learning framework. The chosen backbone is considered EfficientNet-B0.

Pre-processing of Structured Data

Rating Normalization and Personalization A key contribution of this work is the incorporation of a personalized rating transformation procedure that accounts for individual differences in rating behaviour. Since aesthetic judgments are inherently subjective, different participants may systematically use different portions of the rating scale. To mitigate this bias while preserving relative preferences, a user-specific normalization strategy is applied before model training.

Per-User Mean Computation For each participant u , the mean rating across all evaluated interfaces (desktop and mobile) is computed as

$$\mu_u = \frac{1}{N_u} \sum_{i=1}^{N_u} r_{u,i}, \quad (2)$$

where N_u denotes the number of ratings provided by user u .

Global Mean Computation The overall mean rating across all users and interfaces is calculated as

$$\mu_g = \frac{1}{N} \sum_u \sum_i r_{u,i}, \quad (3)$$

where N is the total number of ratings in the dataset.

User-Bias Correction Each raw rating r_{raw} is adjusted by removing the user’s average rating tendency and re-centering it around the global mean:

$$r_{\text{centered}} = r_{\text{raw}} - \mu_u + \mu_g. \quad (4)$$

This transformation compensates for user-specific rating biases while maintaining the overall rating distribution.

Min-Max Normalization The bias-corrected rating is mapped to the interval $[0, 1]$ using min-max scaling:

$$r_{\text{norm}} = \frac{r_{\text{centered}} - 1}{10 - 1} = \frac{r_{\text{centered}} - 1}{9}. \quad (5)$$

Under this transformation, the lowest possible rating (1) corresponds to 0, while the highest possible rating (10) corresponds to 1. Unless otherwise stated, all reported loss values, evaluation metrics, and model predictions in this study are expressed on the normalized $[0, 1]$ scale.

Image Pre-processing To ensure consistency across samples and compatibility with the pretrained backbone network, all user interface (UI) screenshots were subjected to a standardized preprocessing pipeline. Separate transformation strategies were employed for the training and validation datasets. During training, the following transformations were applied sequentially:

1. Resize the image to 224×224 pixels, corresponding to the default input resolution of EfficientNet-B0.
2. Apply random horizontal flipping with a probability of 0.5.
3. Apply random rotation within a range of $\pm 5^\circ$.
4. Apply color jitter with brightness and contrast variation factors of 0.1.

5. Convert the image to a tensor representation.
6. Normalize pixel values using the ImageNet statistics:

The augmentation strategy was intentionally kept mild to preserve the structural integrity of UI layouts while exposing the model to minor visual variations that may arise in real-world screenshots, such as rendering differences and slight orientation changes. Furthermore, ImageNet normalization ensures compatibility with the pretrained EfficientNet-B0 weights and facilitates effective transfer learning.

4.3 Federated Learning Architecture

The model uses the standard EfficientNet-B0 architecture from *torchvision*, which uses *MBConv* (*Mobile Inverted Bottleneck Convolution*) blocks with squeeze and excitation optimization as its main building blocks.

- **Backbone:** The complete EfficientNet-B0 convolutional base (including all MBConv layers and the final conv layer) is loaded with ImageNet-1K pretrained weights. All backbone parameters are fine-tunable.
- **Regression head:** The original 1000-class classification head is removed. In its place, a single fully-connected layer maps the 1280-dimensional feature vector (the output of the backbone’s global average pooling) to one scalar value. No activation function is applied to the output.
- **Output interpretation:** The scalar output represents the predicted aesthetic score on the normalised $[0,1]$ scale. During inference, the prediction is reverse-mapped back to the 1-10 range using the per-group (or per-user) mean and global mean.

The model contains approximately 5.3 million parameters in total. This makes it lightweight, so it’s easy to deploy on devices with limited resources, perfect for a federated learning setting. The architecture is shown in Figure 3.

The training process for each local client (and for the centralised baseline experiments) uses the following, as reported in Table 1.

The same configuration was used for both centralised architecture comparison and federated learning experiments. For the federated scenario, the same setup is replicated on each client, with the global model being grouped via Federated Averaging (FedAvg) after every communication round.

3.2.3 Federated Learning Process

The federated training framework runs on the Federated Averaging (FedAvg) algorithm, with one central server and 5 clients. Each client represents a different demographic group formed by the intersection of gender and occupational status: *female_nonworking*, *female_student*, *female_working*, *male_student*, and *male_working*. At the beginning of the process, the server initialises a global EfficientNet-B0 model with ImageNet-1K pretrained weights. The training proceeds in a series of communication rounds, each comprising three distinct phases: Model Distribution – The server sends the current global

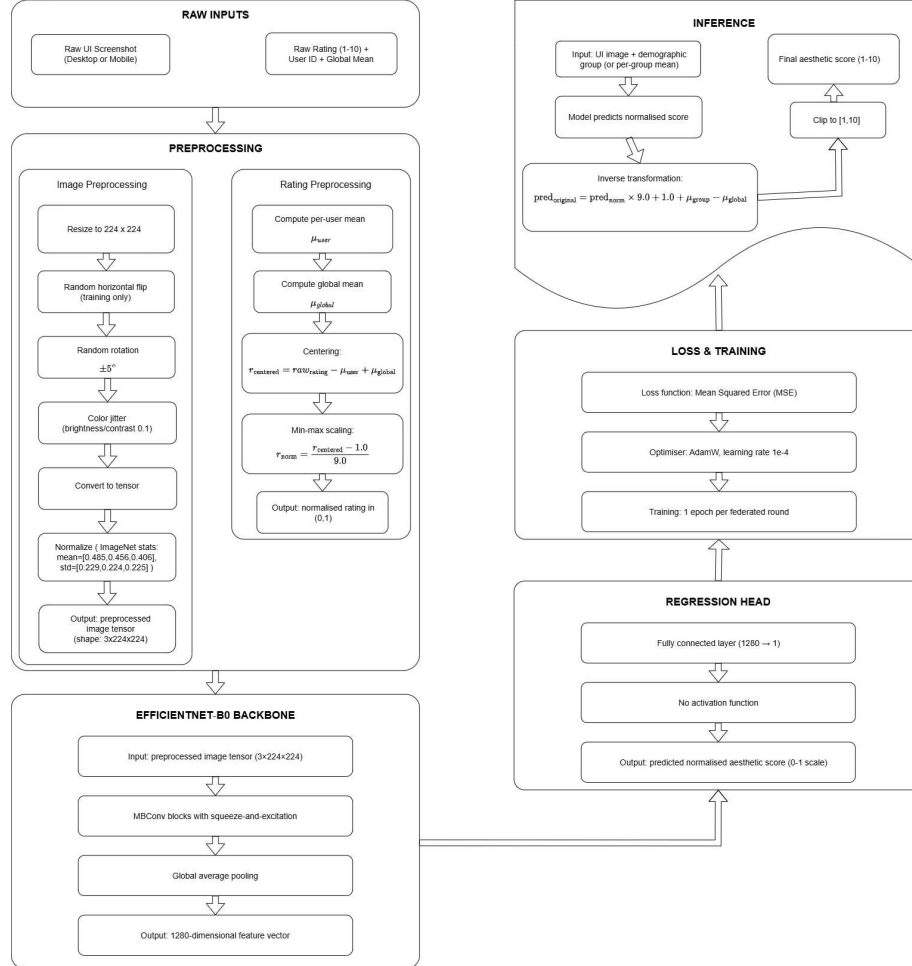


Fig. 3: EfficientNet-B0 Regression Model – Preprocessing, Architecture, and Inference Pipeline

Table 1: Training Configuration and Federated Learning Hyperparameters

Component	Specification
Optimizer	AdamW
Learning Rate	1×10^{-4} (applied uniformly to all parameters)
Batch Size	16
Loss Function	Mean Squared Error (MSE)
Scheduler	ReduceLROnPlateau (mode=min, factor=0.5, patience=3 epochs)
Mixed Precision	Automatic Mixed Precision (AMP) with gradient scaling
Early Stopping	halts if validation loss does not improve for 10 rounds
Number of Communication Rounds	Up to 50
Local Epochs	1 epoch per client per round
Client Weighting	Aggregation weights proportional to the number of training samples per client

model weights to every participating client. **Local Training** – Each client loads the received global model and trains it on its own local dataset for exactly one epoch. The local training uses the same configuration as the centralised baseline: AdamW optimiser with a learning rate of 1×10^{-4} , batch size 16, MSE loss, and Automatic Mixed Precision. After training, the client sends back its updated model weights to the server, and the count of training samples used. The raw data never leaves the client. **Aggregation (FedAvg)** – The server collects all the updated weights and computes a weighted average, where each client’s contribution is weighted proportionally to its number of training samples. This produces a new global model that learns the shared visual knowledge from all demographic groups while handling the uneven data distribution across clients. After aggregation, the updated global model is evaluated on the combined validation sets of all five clients. The validation loss is monitored across rounds, and the learning rate is reduced if no improvement is observed for three consecutive rounds. Training continues for up to 50 rounds, but early stopping is triggered if there is no improvement in validation loss for 10 consecutive rounds. It is important to note that personalisation in this framework is not achieved by maintaining separate local models; rather, it is embedded in the rating transformation. The global model learns to predict a normalised score in the $[0,1]$ range that represents the centered preference of any user. During testing, the score is converted back to the original 1-10 scale using the average rating pattern of a target demographic group (or of a specific user). This design, as shown in Figure 4, allows a single global model to serve diverse users without additional local fine-tuning, keeping the federated system lightweight and easy to deploy.

5 Experimental Results

The **Federated Averaging (FedAvg)** process was run for up to 50 communication rounds, with each of the five demographic clients performing one local epoch per round. The global model was evaluated after every round on the combined validation sets of all clients. Figure 5 shows the global validation MSE

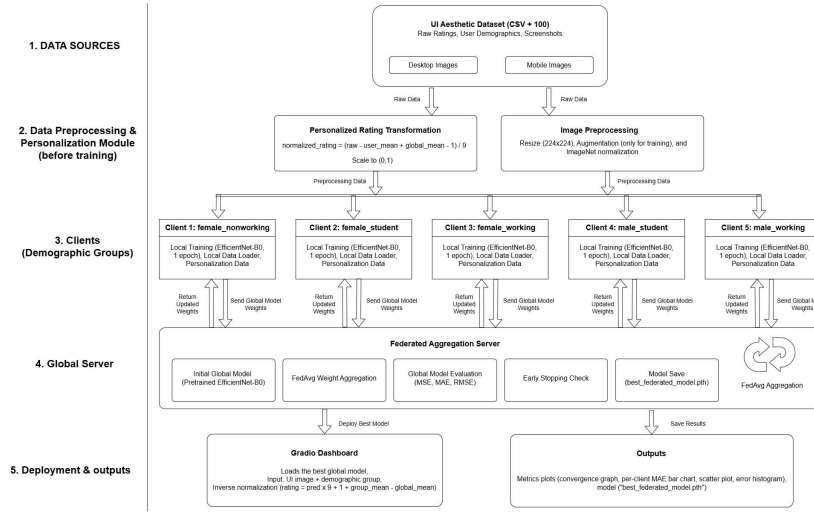


Fig. 4: Federated Learning model for personalized UI aesthetics prediction

as a function of the communication round. The model learned rapidly in the early rounds: the MSE dropped from 0.0702 at round 1 to 0.0583 by round 5, and continued to decrease more slowly thereafter. The lowest validation MSE of 0.0534 was reached at round 33. Because the validation loss did not improve for ten consecutive rounds after round 23, the early-stopping mechanism terminated the process after round 33, and the best checkpoint (round 33) was saved. The final global model therefore achieved $\text{MSE} = 0.0534$, $\text{RMSE} = 0.2312$, and $\text{MAE} = 0.1837$ on the 0 to 1 normalised scale.

The final validation MSE of the global model on each client’s local validation set, taken from the last completed round (round 33). The model performed best on the *female_nonworking* group ($\text{MSE} = 0.0169$) and worst on *female_working* ($\text{MSE} = 0.0890$) as reported in Table 2. This variation is expected: smaller, more homogeneous groups yield lower error, while groups with broader internal preference diversity - such as *female_working* show higher error.

A complementary view is provided by the per-client Mean Absolute Error (MAE) bar chart below. After loading the best global model, the MAE was computed directly on each client’s validation set. The ranking mirrors the MSE table: *female_nonworking* achieves the lowest MAE, while *female_working* and *male_working* show slightly higher MAE, consistent with their larger internal rating variance.

The error distribution histogram further supports this observation, as shown in Figure 6. The errors (predicted minus true) are approximately normally distributed around zero, with no significant skew. The majority of errors fall within ± 0.15 on the normalised scale, which corresponds to roughly ± 1.35 rating points on the original 1–10 scale.

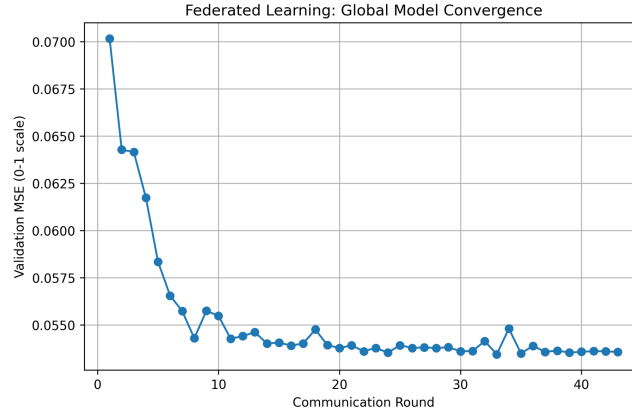


Fig. 5: Convergence Graph

Table 2: Client-wise Validation MAE and User Distribution

Client	Users per Client Ratio	Val MAE
(per 100 users)		
female_nonworking	4	0.0169
female_student	50	0.0551
female_working	4	0.0890
male_student	35	0.0535
male_working	7	0.0783

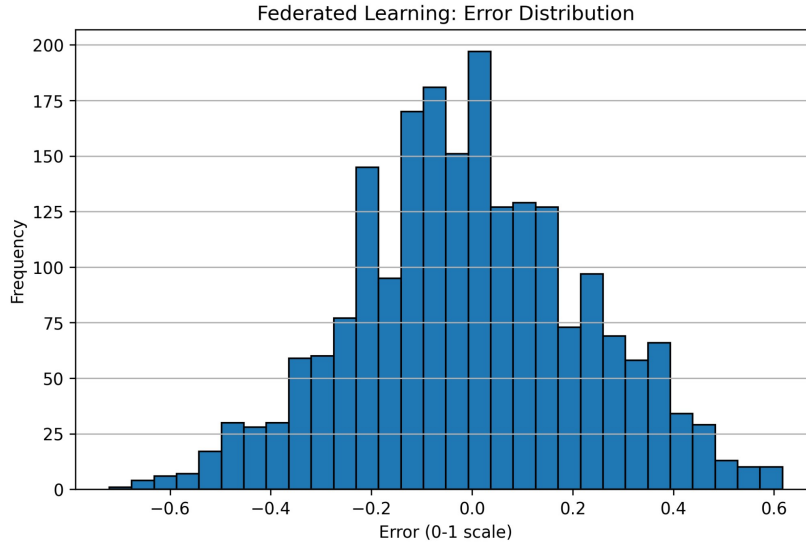


Fig. 6: Error distribution histogram

6 Conclusion and Future Scope

This work presented a privacy-preserving framework for personalized UI aesthetic prediction that combines Federated Averaging (FedAvg) with an EfficientNet - B0 backbone and a novel user-adaptive rating normalisation. Instead of centralising sensitive user ratings, the system distributes the learning process across five demographic client groups. Each client trains the shared model locally on its own data, and only the updated model weights are sent to a central server for aggregation. The raw ratings never leave the local device. Federated Averaging reached its goal in 33 rounds of communication, ending up with a global validation MSE of 0.0534 on the normalized scale. So, no loss in accuracy compared to the centralized baseline where everything's pooled together. Even with everyone's preferences scattered and unique, the system combined them without sharing anyone's raw ratings. Personalisation was realised through a user-mean centering and min-max scaling pipeline. By removing individual rating biases before training and re-injecting them via the corresponding demographic group mean at inference, a single global model was able to serve users from diverse backgrounds. The per-client analysis confirmed that the model adapts to different demographic groups, with MAE varying in line with the internal homogeneity of each group.

The most prominent limitation is the global model's inclination to predict scores concentrated around the dataset's average (approximately 0.5 - 0.6 on the normalised 0-to-1 scale). The current system deploys a single shared global model to all clients. Personalisation is achieved exclusively through the rating transformation pipeline (user-mean centering and demographic group bias). No client-specific fine-tuning or local head personalisation is performed after the federated training.

References

1. R. Maity and S. Bhattacharya, "Is my interface beautiful?—a computational model-based approach," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 1, pp. 149–161, 2019.
2. G. Lindgaard, G. Fernandes, C. Dudek, and J. Brown, "Attention web designers: You have 50 milliseconds to make a good first impression!" *Behaviour & information technology*, vol. 25, no. 2, pp. 115–126, 2006.
3. C. Yang, B. Yuan, K. Wu, and J. Ye, "Aesthetic evaluation method for interactive interface layouts on the basis of visual cognitive multi-attribute fusion decision making," *International Journal of Industrial Ergonomics*, vol. 105, p. 103673, 2025.
4. D. Ceneda, C. Collins, M. El-Assady, S. Miksch, C. Tominski, and A. Arleo, "A heuristic approach for dual expert/end-user evaluation of guidance in visual analytics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 30, no. 1, pp. 997–1007, 2023.
5. R. Maity, A. Uttav, G. Verma, and S. Bhattacharya, "A non-linear regression model to predict aesthetic ratings of on-screen images," in *Proceedings of the annual meeting of the australian special interest group for computer human interaction*, 2015, pp. 44–52.

6. R. Maity and S. Bhattacharya, "A model to compute webpage aesthetics quality based on wireframe geometry," in *IFIP Conference on Human-Computer Interaction*. Springer, 2017, pp. 85–94.
7. —, "Relating aesthetics of the gui text elements with readability using font family," in *Proceedings of the 2018 ACM Companion International Conference on Interactive Surfaces and Spaces*, 2018, pp. 63–68.
8. M. Daryanavard Chouchenani, A. Shahbahrani, R. Hassanpour, and G. Gaydadjiev, "Deep learning based image aesthetic quality assessment-a review," *ACM Computing Surveys*, vol. 57, no. 7, pp. 1–36, 2025.
9. Z. Zheng, D. Yang, L. Zeng, and N. Mughees, "A deep learning framework for objective aesthetic evaluation of indoor landscapes using cnn-gnn model," *Scientific Reports*, vol. 15, no. 1, p. 40810, 2025.
10. S. Saha, D. Basumatary, A. Senapati, and R. Maity, "Is there any further scope for improving the efficiency of modern websites?" in *2021 6th International Conference for Convergence in Technology (I2CT)*. IEEE, 2021, pp. 1–7.
11. S. Cao, N. Ma, J. Li, X. Li, L. Shao, K. Zhu, Y. Zhou, Y. Pu, J. Wu, J. Wang et al., "Artimuse: Fine-grained image aesthetics assessment with joint scoring and expert-level understanding," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026, pp. 15 313–15 322.
12. C. Redies, R. Bartho, L. Koßmann, B. Spehar, R. Hübner, J. Wagemans, and G. U. Hayn-Leichsenring, "A toolbox for calculating quantitative image properties in aesthetics research," *Behavior Research Methods*, vol. 57, no. 4, p. 117, 2025.
13. R. Maity, A. Madrosiya, and S. Bhattacharya, "A computational model to predict aesthetic quality of text elements of gui," *Procedia Computer Science*, vol. 84, pp. 152–159, 2016.
14. Y. Tian and S. Lee, "Aesthetic evaluation of visual perception in ai-generated short films via integrated support vector machine and attention mechanism," *IEEE Access*, 2025.
15. R. Maity and S. Bhattacharya, "A quantitative approach to measure webpage aesthetics," *International Journal of Technology and Human Interaction (IJTHI)*, vol. 16, no. 2, pp. 53–68, 2020.
16. N. Murray, L. Marchesotti, and F. Perronnin, "Ava: A large-scale database for aesthetic visual analysis," in *2012 IEEE conference on computer vision and pattern recognition*. IEEE, 2012, pp. 2408–2415.
17. Y. Zhou, "The application of deep network and reinforcement learning for art design in graphical user interface wireframe generation," *Automatic Control and Computer Sciences*, vol. 59, no. 3, pp. 402–415, 2025.
18. R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Studying aesthetics in photographic images using a computational approach," in *European conference on computer vision*. Springer, 2006, pp. 288–301.
19. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
20. S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "Predicting image aesthetics with deep learning," in *International Conference on advanced concepts for intelligent vision systems*. Springer, 2016, pp. 117–125.
21. J. McCormack and A. Lomas, "Deep learning of individual aesthetics," *Neural Computing and Applications*, vol. 33, no. 1, pp. 3–17, 2021.
22. B. Koonce, "Efficientnet," in *Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization*. Springer, 2021, pp. 109–123.