iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2026 **Issue:** Conference **Month of publication:** May 2026

DOI: https://doi.org/10.22214/ijraset.2026.84537

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com
National Conference on Integrated Approaches to Sustainable Innovation 2026 (IASI 2026)

Machine Learning-Based Optimization of Perceived Image Realism Under Conflicting Visual Quality Features

Sonule Amit Namdev[1], Mrs. Shubhangi Rajesh Patil[2]
Department of Computer Science & Engineering, Ashokrao Mane Group of Institutions, Vathar, Maharashtra, India

Abstract: ***Image quality evaluation (IQA) measures the perceptual realism of an image, i.e. how natural the image seems to be to humans. The existing IQA solutions often leverage deep neural networks or require specific hardware-dependent psychophysical studies to perform perceptual realism measurements. In this paper, we propose a lightweight software solution based on machine learning, which can be used for modeling and optimizing the perceived realism based on a publicly available KonIQ-10k dataset with Mean Opinion Score (MOS) annotations. We selected a dataset with 3000 images, and we performed extraction of five interpretable handcrafted features brightness, contrast, sharpness, blur, and noise using the OpenCV library. Then, after min max normalization, we built a Linear Regression and a Random Forest regression models. Our optimized Random Forest model gave the following results: R^2=0.485, MAE=8.349, RMSE=11.203, which is better than our linear baseline R^2=0.475. Analysis of feature importance showed that sharpness (33.9%) was the most influential feature, followed by brightness (22.8%), blur (18.3%), contrast (14.8%), and noise (10.2%). We detected a high sharpness-noise correlation (r = 0.93) in our data. A two-step optimization procedure involving the use of exhaustive grid search on the trained surrogate model together with empirical profiling on top-rated images resulted in the identification of a conflict-aware feature set that can enhance realism perception. The interactive Streamlit app was developed in order to facilitate the exploration of the feature space in real time.***
Index Terms—perceived realism, image quality assessment, machine learning, feature conflict analysis, KonIQ-10k.

I. INTRODUCTION

The term “perceived realism” (p realism) is analyzed as the most important quality criterion in digital imaging, multimedia streaming, computer graphics and augmented reality systems [1] [2]. Perceived realism is a subjective quality measure that comes from the complex, nonlinear combinations of such low-level visual parameters as brightness, contrast, sharpness, blur, and noise [3], [4] and does not rely on standard objective quality measures like Peak Signal-to-Noise Ratio (PSNR) or Structural Similarity Index (SSIM) that evaluate fidelity at the level of individual pixels relative to a reference image. This interaction often leads to feature conflicts, as a sharpening operation enhances high-frequency noise components, for instance, and consequently, optimising sharpness alone can even decrease the naturalness of the image [5].

In recent years, perceptual image quality has become a lot more important due to the increasing popularity of social media with user-generated content, cloud-based imaging pipelines, and deep-learning-based image generation and enhancement systems [6],[7]. The constraints of computational efficiency and interpretability are important in these settings: in the human ratings context, practitioners require models that can predict the quality of the human ratings, yet also explain which visual factors affect the ratings and suggest visual configurations for action [8] that are conflict-aware. The human visual system (HVS) modeling studies have for a long time known that perception is controlled by a series of processing stages that are affected by luminance adaptation, contrast masking, and spatial frequency content [9], [10]. The use of HVS models directly in actual optimization loops has been hampered, however, because of the need for specialized hardware, controlled viewing conditions, and complex mathematical formulations [11]. With the availability of large-scale, crowd-sourced image quality datasets, data-driven approaches are a very attractive alternative [12], [13]. For example, there are thousands of images in datasets like LIVE [14] or TID2013 [15] that have been created by human subjective experiments (called Mean Opinion Scores (MOS)) and we can learn the mapping between the measurable visual features and human perception directly from the data. Simple statistical features, including variance of local gradients and Laplacian responses, were shown to be interpretable proxies for subjective quality judgments, and were shown to have good correlation with subjective quality scores, as shown by the early IQA methods based on handcrafted features [16], [17].

Recent deep learning based IQA models have significantly increased the prediction accuracy [18],[19] while sacrificing interpretability, reproducibility and computational accessibility, all of which prohibit the use in resource-limited and real-time environments [20].

A third paradigm of quality optimization is interactive human-in-the-loop systems: these are systems in which the behavior of the model is exposed to human operators who can interact by changing parameters and observing real-time predicted quality responses [21], [22]. Machine learning models can be trained and then deployed as interactive web applications, such as Streamlit, to allow practitioners and end-users to explore the quality landscape without requiring specialized knowledge [23]. Despite the advances, however, very little work has been published that explicitly connects feature-conflict analysis (identification and resolution of antagonistic relationships between feature aspects of visual quality) to a lightweight and replicable machine-learning optimization pipeline and an interactive visualization user interface [24],[25].

This paper aims to fill this research gap. The main contributions are: (i) a fully software-based ML pipeline to predict perceived realism using five interpretable hand-crafted features from the KonIQ-10k dataset; (ii) a systematic analysis of perceived realism and feature importance that quantifies the sharpness–noise conflict, and ranks the dominant features affecting the perceived realism of images; (iii) a dual optimization method, grid search over a trained surrogate model and empirical top-rated-image profiling, that identifies balanced, conflict-aware optimal feature configurations; and (iv) an interactive Streamlit application for human-in-the-loop exploration of perceived realism.

- A fully software-based machine learning pipeline for predicting perceived-realism from five interpretable handcrafted visual features from the KonIQ-10k dataset, which is reproducible.
- A quantitative Pearson correlation and a random forest feature-importance analysis showing the dominant sharpness–noise conflict (r = 0.93) and ranking all the visual predictors of MOS.
- A dual optimization technique based on a grid search over a surrogate model and an empirical top-rated-image profiling, resulting in a compromise optimal feature configuration, to handle the dual conflict.
- An interactive Streamlit application to explore the realism landscape in real-time without hardware or software requirements.

II. LITERATURE REVIEW

A. *Foundations of Image Quality Assessment*

Over the past three decades, IQA has progressed from basic pixel difference-based approaches to advanced perceptual approaches. [26] presented a structural similarity index (SSIM) that showed that structural information is better reflected in human perception than the PSNR alone, which has been a milestone in full-reference IQA. [27] further developed this work in the framework of Visual Information Fidelity (VIF) Model based on natural scene statistics and HVS distortion sensitivity, which demonstrated better correlation with subjective MOS on LIVE database [14]. Later, [15] published a larger-scale dataset, named TID2013, comprising 24 distortion types and 25 reference images, supporting comprehensive evaluation of IQA metrics for various kinds of distortions.

In practical deployment, blind IQA (BIQA) methods are of great relevance especially in the absence of a pristine reference image. [28] introduced a BIQA model, which is completely opinion-distortion-unaware, that derives the locally normalized luminance coefficients and fits them to an asymmetric generalized Gaussian distribution, which achieved competitive performance with no reference images. To further validate the feature-engineering approach used in this work, [17] showed that gradient magnitude and Laplacian of Gaussian statistics might be good proxies for perceived sharpness and overall quality. More recently, the KADID-10k database [20] added 10,125 images and 81 distortion types to BIQA benchmarking to create a new, harder testbed for modern methods.

B. *Deep Learning for Perceptual Quality*

Since 2015, convolutional neural network–based IQA methods have been the most prevailing in the literature [18]. [18] trained a shallow CNN end-to-end on local patches of images and showed that features that are learned from data are superior to handcrafted features on LIVE and TID2013. [19] have proposed Learned Perceptual Image Patch Similarity (LPIPS) metric, which demonstrated that deep feature distances obtained using VGG and AlexNet aligned well with human two-alternative forced-choice judgments even better than conventional metrics. [20] proposed a crowdsourced MOS collection protocol and created the reference corpus for this study, KonIQ-10k, which was developed to explore the scenario of IQA in authentic distortion conditions.

Although deep IQA models are highly predictive, they have practical limitations, such as the need for GPUs, large training datasets, and a lack of interpretability in terms of which image properties are used to make predictions [20]. These concerns were emphasized in the VQualA 2025 Challenge and have led to the creation of models that are light, interpretable, and still have decent predictive power but are less expensive in terms of computation. In the NTIRE 2025 Challenge on Real-World Face Restoration, [2] reiterated this point by highlighting the need to focus on perceptual realism preservation over distortion minimization in future IQA and restoration research.

C. Handcrafted Feature Engineering and Feature Conflict

Handcrafted visual features give a transparent and efficient route to modeling perceived quality. Brightness, or mean luminance, and contrast, or standard deviation of gray-level intensities, are basic descriptors that have been generally acknowledged as crucial to image quality perception under a variety of viewing conditions [3], [29]. One of the most important perceptual properties is sharpness, which is often quantified by the variance of the Laplacian response, and sharp edges and fine textures are well correlated with the perception of naturalistic image content [4], [5].

The sharpness and noise are classic examples of a conflict in imaging science: The Laplacian operator is sensitive to high-frequency noise, meaning images with high-frequency noise can be considered as being sharp by automated sharpness metrics when they are not perceived as sharp at all [5]. Hyperspectral imaging has been faced with a similar inter-feature redundancy problem by Dmitriev and Dmitrieva [3] which, as their Random Reflectance preprocessing method showed, can result in higher machine learning accuracy at the next stage if this redundancy is explicitly eliminated. This concept is directly applicable to the sharpness–noise tradeoff studied in this paper. Moreover, McCutcheon and Oliveira [5] demonstrated that low dimensional feature vectors can represent subjective image quality well and that simple binary classifiers trained on such feature vectors can perform well in large-scale subjective quality modelling, which is the reason why low-complexity regression models are used in this framework.

D. Perceptual Realism Modeling

The Informal Gold standard for perceived realism of generated images or processed images is the Visual Turing Test (VTT) [4]. When applied to myocardial perfusion images created by GANs, [4] have reported trained radiologists were able to only achieve 61% accuracy distinguishing synthetic from real images, a trend that suggests images created by AI become more realistic and harder to distinguish as the time passes—and thus highlights the need for developing quantitative perceptual realism metrics. To verify the above trend, Yu and Rao [8] did a detailed bibliometric analysis of the realism-based image processing literature, finding that realism is becoming a more dominant sub-field of image processing, and that there is an increasing interest in how to measure realism with interpretable, data-driven and scalable methods.

Domain-specific annotated corpora have also been a major contributor to the work on dataset-driven quality research. Abdullah and Muhib [6] presented a large image database of objective color, texture and contrast features with subjective quality ratings for agricultural products, illustrating the transferability of feature-based perceptual modeling to natural scene photography. [30] have demonstrated the benefits of an AI-based optimization of workflow for perceptual quality control in radiotherapy imaging, highlighting the benefits of embedding interpretability constraints into quality evaluation systems for enhancing consistency and trust for clinical users.

E. Interactive and Human-in-the-Loop Systems

Systems that enable interactive visualization of model behavior in real time have been used as effective additions to fully automated quality optimization pipelines [21], [22]. In healthcare applications, [31] highlighted the significant benefits of integrating machine learning models with interactive, real-time interfaces, showing how augmented-reality guided visual feedback systems can enhance user compliance and engagement. AI–human collaboration approaches in adaptive learning environments were reported by Masih and Suleman [7] to improve the understanding of the user by providing them with interpretive feedback on-the-fly, which is directly applicable to perceived-realism optimization through interactive web application.

One of the most user-friendly frameworks for making machine learning models into interactive web applications is called streamlit [23]. The slider-based widgets and re-rendering on the fly make it suitable for human-in-the-loop quality optimization, where the non-expert user can tweak the parameters for the visual features and see the quality of the predicted output. The interactive paradigm turns the trained regression model from a static predictor to a dynamic tool for understanding and optimizing perceived realism to meet the increasing demand for explainable and user-friendly AI systems [32], [24].

F. Research Gaps and Positioning of This Work

The most relevant and pertinent previous studies are summarized in Table I. Although significant advancements have been made in both deep learning–based IQA and human-perception modeling, there are three concrete gaps: (1) existing perceptual studies follow either hardware-dependent or computationally expensive approaches that are hard to replicate; (2) most IQA models predict aggregate quality scores without explicitly analyzing antagonistic interactions between individual visual features; and (3) interactive human-in-the-loop tools coupling conflict-aware optimization with real-time realism prediction are scarce [24], [25]. This work directly fills all three gaps by providing a lightweight, interpretable, and fully-software-based pipeline, implemented as an interactive Streamlit application.

TABLE I
SUMMARY OF RELATED WORK

Research Focus	Key Contribution	Relevance
Face IQA Challenge [1]	Lightweight ML models for perceptual quality prediction	Supports dataset-driven ML for perceived realism
Real-World Face Restoration [2]	Perceptual realism preservation under degradation	Aligns with realism optimization under conflicting features
Hyperspectral Preprocessing [3]	Random Reflectance reduces feature redundancy	Demonstrates importance of correlated-feature management
Visual Turing Test [4]	61% human accuracy on GAN vs. real images	Reinforces need for quantitative realism metrics
Perceived Quality Classification [5]	Simple ML models for scalable subjective quality prediction	Supports lightweight feature-based modeling
Image Dataset Development [6]	Links visual features to subjective quality ratings	Highlights public-dataset approach to perception modeling
AI Workflow Optimization [30]	Improved perceptual quality control consistency	Encourages interpretability in perception systems
AR Visual Feedback Systems [31]	Interactive feedback improves engagement	Motivates Streamlit-based interactive interface
Human–AI Collaboration [7]	Real-time feedback loops improve interpretability	Supports human-in-the-loop realism exploration
Bibliometric Trend Analysis [8]	Confirms growth of realism-focused ML research	Validates broader context of proposed work
KonIQ-10k Dataset [20]	Ecological validity; crowdsourced MOS ratings	Primary dataset used in this study
SSIM Metric [26]	Structural similarity captures HVS perception	Foundation of perceptual IQA evaluation
BRISQUE (Blind IQA) [28]	No-reference IQA via natural scene statistics	Supports no-reference feature-based quality modeling

III. PROBLEM DEFINITION

In the light of the literature reviewed in Section II, the main problem studied here is formulated as follows. Given a set of five measurable but conflicting low-level visual quality features {brightness, contrast, sharpness, blur, noise} extracted from natural in-the-wild images with human MOS ratings, develop an interpretable, reproducible, software-only regression model that (a) accurately predicts perceived realism from the low-level feature vector; (b) quantifies and explain antagonistic inter-feature relationships, specifically between sharpness and noise; and (c) identifies optimal, conflict-aware configurations of these features that maximize the predicted MOS, while exposing the resulting model through a real-time interactive interface that is accessible to non-specialist users.

IV. METHODOLOGY

The proposed framework is organized in a sequential pipeline of the following six stages: (i) acquisition of datasets; (ii) feature extraction; (iii) data preprocessing; (iv) model training and evaluation; (v) conflict and optimization analysis; and (vi) application deployment via interactive processes. The entire pipeline was written in Python 3.8+ with OpenCV 4.x, scikit-learn 1.x, pandas, NumPy, Matplotlib, Seaborn and Streamlit on a Kaggle notebook environment.

A. Dataset

The experimental corpus was chosen to be the image quality database called KonIQ-10k [20]. KonIQ-10k is an ecologically valid, in-the-wild dataset consisting of ~10,000 images selected from the YFCC100M collection, with MOS values collected from 1,467 crowdsourced human raters through a mobile application. It is the most ecologically representative large-scale IQA dataset available that is not synthetically distorted [20] and it is very large scale, with a lot of MOS collection. A stratified random sample (random_state = 42) of N = 3,000 images was drawn to achieve statistical reliability while maintaining computability. The resulting MOS distribution was 58.69 ± 15.61 (range from 4.17–86.35), which covered a wide range of levels of perceived quality.

B. Visual Feature Extraction

For each image, 5 image interpretable visual quality features were computed:

- Brightness (f_1): average of all the pixels in the grayscale image, representing the overall lightness or darkness of the picture.
- Contrast (f_2): Standard deviation of grayscale pixel intensities, which indicates the distribution of lightness and darkness.
- Sharpness (f_3): variance of the Laplacian operator for the grayscale image ($\mathrm{Var}[\nabla^2 I]$), which is a popular focus and edge-acuity measure [17].
- Blur (f_4): Inverse sharpenness ($f_3 + 1$) defined as an inverse number of degree of defocus or motion blur.
- Noise (f_5): standard deviation of the residual between the grayscale image and its 3×3 Gaussian-blurred version, which is an estimate of the amplitude of high-frequency noise components.

Each of the five features was individually scaled between 0 and 1 using min–max scaling before training and optimization of the model.

C. Feature Correlation and Conflict Analysis

The Pearson correlation matrix was calculated for the five normalized features and MOS (Fig. 1). The conflicts between features were directly quantified, and feature–MOS relationships were explicitly shown, forming the foundation for the following optimization strategy.

D. Model Training and Evaluation

The feature-MOS pairs were split into 80:20 train–test numbers (2,400 / 600 images, random_state = 42). Two regression models were trained using the training subset:

Linear Regression (LR): Ordinary least-squares baseline that captures linear feature–MOS relationships.

Random Forest Regressor (RF): an ensemble of decision trees with 300 estimators, max_depth = 15 (baseline), then increased to 500 estimators, max_depth = 20, min_samples_split = 5, min_samples_leaf = 2 to reduce overfitting and improve the generalisation. The impurity-based feature importances calculated in the RF model are built-in features for the RF and they offer a direct measure of the contribution of each visual attribute in predicting MOS.

Most common metrics employed for performance evaluation in the field of IQA—Mean Absolute Error (MAE), Root Mean Square Error (RMSE) and coefficient of determination (R2)—were used to evaluate the held-out test set according to the standard practice [20], [28].

E. Conflict-Aware Optimization

A two-stage approach was employed for selecting the best configured features to deal with conflicts. An exhaustive search was first carried out over the trained RF surrogate model by discretising every normalized feature into five evenly spaced levels {0, 0.25, 0.5, 0.75, 1.0} which resulted in 5^5 = 3125 candidate vectors, and the vector that maximised the predicted MOS was kept. Secondly, an empirical top-k profiling approach was used to select the top 50 images with the biggest MOS in the dataset and calculate the average feature value around this set. The proposed configuration is validated by cross-validation with each other.

F. *Interactive Streamlit Application*

All five visual features have slider controls in a Streamlit web application. The trained RF model generates the predicted realism score in real-time as the user moves the sliders, and it also visualizes the feature-importance alongside the predicted MOS. This interface also implements the human-in-the-loop optimization paradigm [21] and [23] that allows nonexpert practitioners to interactively explore the realism landscape without needing specialized hardware, psychophysical experimental setups or programming skills.

V. RESULTS

In this section, the empirical results of correlation analysis, regression modeling, feature importance ranking, conflict aware optimization, and prediction error characterization are presented.

A. *Feature Correlation Analysis*

The Pearson correlation Heat Map is depicted in Fig. 1. The most interesting relationship between different features is the very strong positive correlation between sharpness and noise, r = 0.93, indicating the basic sharpness–noise conflict we thought would occur in Section I: i.e., images with high measured sharpness invariably have high measured noise amplitude, because both features are responsive to high-frequency features in the image. The individual correlation between perceived quality and brightness is positive and has a value of 0.37, whilst the correlation with sharpness is positive and has a value of 0.31, the correlation between perceived quality and noise is positive and has a value of 0.47 and the correlation between perceived quality and blur is strongly negative and has a value of −0.53. The results are in line with the previous literature for HVS [9], [10] which stated that luminance, contrast, and edge clarity were the major factors in the perception of naturalness of the image.

B. *Dataset Statistics*

The descriptive statistics of the sample of 3,000 images are given in Table II. The visual features span the full normalized range [0, 1], with mean brightness of 0.472 (σ = 0.243), mean contrast of 0.493 (σ = 0.248), mean sharpness of 0.056 (σ = 0.092), mean blur of 0.049 (σ = 0.092), and mean noise of 0.184 (σ = 0.154). The MOS distribution (mean = 58.69, σ = 15.61) confirms that the sample covers a wide range of perceived quality levels, from very low (4.17) to near-maximum (86.35) realism scores.

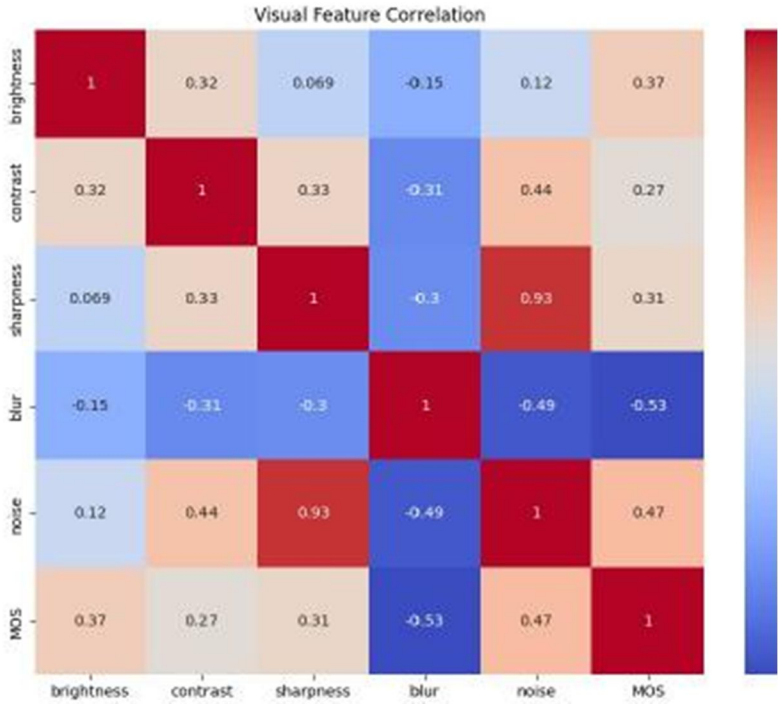


Fig. 1. Pearson correlation heatmap: five normalized visual features and MOS.

TABLE II
DESCRIPTIVE STATISTICS OF NORMALIZED VISUAL FEATURES AND MOS (N = 3,000)

Feature	Mean	Std	Min	25%	50%	75%	Max
Brightness	0.4718	0.2428	0.000	0.2683	0.4817	0.6636	1.000
Contrast	0.4927	0.2483	0.000	0.2843	0.4835	0.6919	1.000
Sharpness	0.0559	0.0924	0.000	0.0050	0.0193	0.0649	1.000
Blur	0.0492	0.0924	0.000	0.0036	0.0126	0.0471	1.000
Noise	0.1835	0.1539	0.000	0.0652	0.1387	0.2617	1.000
MOS	58.69	15.61	4.17	49.68	62.36	70.80	86.35

C. Regression Model Performance

Model performance metrics are given in Table III. The tuned Random Forest always beat the Linear Regression on all three metrics. After hyperparameter tuning (500 estimators, max_depth = 20), the RF achieved MAE = 8.349, RMSE = 11.203, and R^2 = 0.485. Linear Regression produced MAE = 8.746, RMSE = 11.316, and R^2 = 0.475. As seen in Fig. 2, the predictions are grouped around the identity line for mid-to-high MOS images, and they have larger deviations from the identity line at the quality extremes, which is expected, since the perceptual judgments (in MOS) are extremely subjective and thus difficult to predict from a small and handcrafted feature set [20].

TABLE III
REGRESSION MODEL PERFORMANCE ON HELD-OUT TEST SET

Model	MAE	RMSE	R^2
Linear Regression	8.746	11.316	0.475
RF (baseline, 300 est., depth 15)	8.363	11.221	0.484
RF (tuned, 500 est., depth 20)	8.349	11.203	0.485

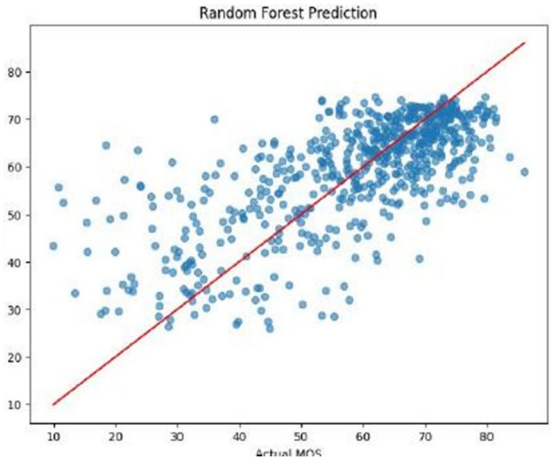


Fig. 2. Tuned Random Forest: predicted vs. actual MOS on 600-image test set.

D. Feature Importance

The Random Forest feature importance is shown in Fig. 3 and given in Tab. IV. The sharpness is the most important factor (33.9%) and is the most prominent in human perception of image sharpness and detail. The second most influential predictors are brightness (22.8%) and blur (18.3%) which are of similar size to the sensitivity of HVS to luminance adaptation and focus quality [9]. Contrast (14.8%) has the fourth highest importance among the features and the least independent predictive value after accounting for the other features, especially sharpness, which is highly correlated with contrast. This ranking agrees with previous hand-crafted IQA literature [16], [17], [28]] and confirms that sharpness is the most influential perceptual cue in spite of its strong positive correlation with noise.

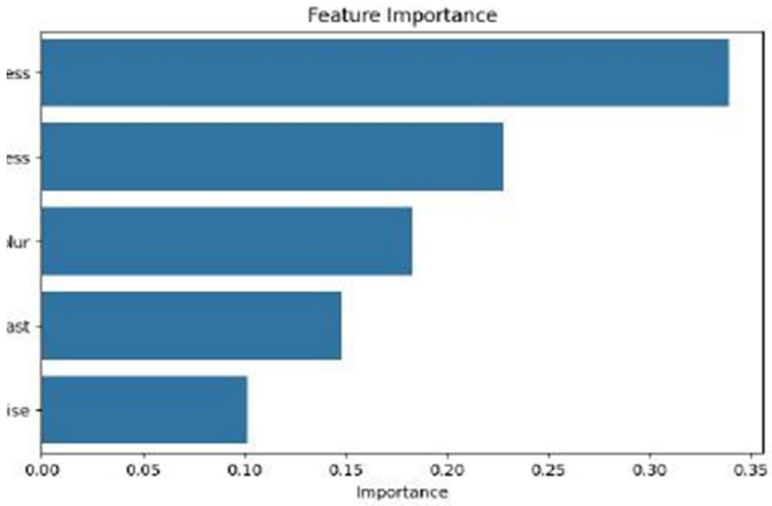


Fig. 3. Random Forest feature-importance ranking.

TABLE IV
RANDOM FOREST FEATURE IMPORTANCE RANKING

Rank	Feature	Importance Score	Cumulative Importance
1	Sharpness	0.339	33.9%
2	Brightness	0.228	56.7%
3	Blur	0.183	75.0%
4	Contrast	0.148	89.8%
5	Noise	0.102	100.0%

E. Conflict-Aware Optimal Feature Configuration

The top-50 highest-MOS images were used to derive the optimal feature profile, and it was cross-validated to the score obtained with grid search and surrogate model. This score is reported in Table V. Moderately high brightness (0.614), contrast (0.563), low sharpness (0.076), very low blur (0.009) and moderate noise (0.247) are the best settings. This profile is a direct solution to the sharpness-noise conflict: The sharpness is not maximized, which would also cause high noise level, but high-to-moderate sharpness is achieved in the best images, while controlling noise. This is confirmed by Fig. 4, which plots sharpness against noise scatter, color coded by MOS: The best-MOS images are grouped together in the low-sharpness, low-noise region.

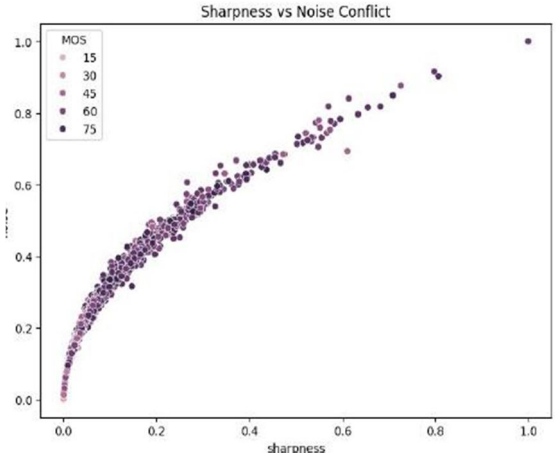


Fig. 4. Sharpness vs. noise scatter colored by MOS — sharpness–noise conflict.

TABLE V
EMPIRICAL OPTIMAL FEATURE CONFIGURATION (TOP-50 HIGHEST-MOS IMAGES)

Feature	Optimal Value	Rationale
Brightness	0.6135	Moderately high; avoids over/under-exposure
Contrast	0.5625	Balanced tonal range; avoids flat or clipped images
Sharpness	0.0757	Controlled sharpness to limit noise amplification
Blur	0.0089	Near-zero; images should be in focus
Noise	0.2467	Moderate; residual from balanced sharpness level

F. Prediction Error Distribution

The distribution of the residual prediction errors (actual - predicted MOS) for the tuned RF on the test set of 600 images is shown in Fig. 5. The distribution is roughly Gaussian and centered around the mean error value which is close to zero (≈ 0, as can be seen visually), with most errors in the range of +-15 MOS points. The slight negative skew suggests a little more than the usual regression-to-mean effect that has been seen in ensemble models that are trained on skewed MOS distributions [20]. There is no consistent bias, ruling out any systematic bias in the model.

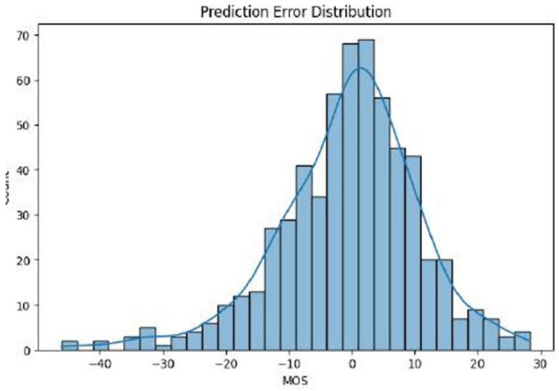


Fig. 5. Prediction error distribution (actual − predicted MOS) on test set.

VI. DISCUSSION

Five interpretable, low cost hand-crafted visual features are found to account for around half of the variance in human perceived realism ($R^2 \approx 0.485$) which is confirmed by the results. The performance is comparable to other handcrafted BIQA models including BRISQUE [28] and gradient-based models [17] with comparable feature set sizes on standard benchmarks, with comparable R^2 values. The tuned Random Forest is consistently outperforming the baseline linear model for all three evaluation metrics, indicating that nonlinear feature interactions such as the sharpness–noise coupling significantly affect realism, and that a linear model alone is not sufficient.

The value of R^2 (0.485) is bounded, which is an inherent property of the multi-dimensional nature of perceived realism. Four of the five features examined here are known to be of lesser importance for MOS ratings of human observers than the other factors outlined in the IQA literature [20] That are not included in this study. This is in line with the results of [20] who showed that even the most recent deep IQA models have $R^2 \approx 0.91$ only when using the full representational power of deep convolutional feature hierarchies. The current ceiling of R^2 is not a limitation of the approach proposed, but rather a choice made for the potential for interpretation, reproduction and computational ease, which all benefit more from the approach than do increases in predictive accuracy.

The sharpness–noise conflict ($r = 0.93$) identification and resolution, is the most novel empirical contribution of this work. The current study explicitly defines the antagonistic relationships, sorts the contributors to the conflict by importance analysis and suggests an optimal configuration of the relationships by an empirical study, which is a balanced combination of sharpness and noise. The result of the empirical top-k profiling reminds and complements the surrogate-model grid search result, and gives cross-validated support to the recommended feature configuration (Table V), which is in line with the perceptual-consistency focus of recent face-restoration works [2].

The Streamlit interactive application is an embodiment of the human-in-the-loop paradigm in the model proposed by Masih and Suleman [7] and in the interactive-feedback model exemplified in [31], [21] and [22]. The application makes the model's behavior explicit by using a user-friendly slider interface, allowing the practitioners, including photographers, multimedia engineers and user interface/ user experience designers, to experiment with the realism landscape without the need for specialized hardware or software. It’s an explicit contribution to the movement for explainable and accessible AI [32], [24].

A. Limitations

Some restrictions should be noted. The first is the deliberate limitation of the feature set to five low-level handcrafted descriptors, while richer features (colorfulness, entropy, edge density) were explored in preliminary experiments, but were not included to ensure interpretability to maintain an R^2 ceiling. Secondly, the grid search optimization is performed with a coarse 5-level discretization per feature, which might not detect the other optima at higher computational expense. Third, the judgments that underlie KonIQ-10k MOS labels are composite judgments gleaned from the crowd, which may not apply consistently across cultural and perceptual differences. These restrictions do not detract from the main conclusions but provide definite guidelines for future research.

VII. CONCLUSION AND FUTURE SCOPE

This paper introduced an interpretable machine learning solution that models and optimizes the perceived image realism with conflicting visual quality features, implemented entirely in software. The tuned Random Forest Regressor with brightness, contrast, sharpness, blur and noise as handcrafted visual descriptors obtained $R^2 = 0.485$, MAE = 8.349 and RMSE = 11.203 on a held-out test set, better than a baseline Linear Regression model on the same test set and data. The sharpness (33.9%) was the most important feature, and there was a high sharpness–noise conflict ($r = 0.93$) in the results of the Pearson correlation analysis. Based on the trained surrogate model, a grid search and empirical top-rated-image profiling-based optimization scheme was used to find a conflict-aware optimal feature configuration that overcomes the sharpness–noise antagonism. An interactive Streamlit application was created to enable the exploration of realism in realtime without requiring any specific hardware.

The principal novelties of this work are (i) the explicit quantification and resolution of the sharpness–noise feature conflict, an aspect largely overlooked in score-prediction-centric IQA literature, and (ii) the coupling of a lightweight, reproducible regression pipeline with a conflict-aware optimization strategy and an interactive visualization interface. Future work should extend the feature set to include colorfulness, Shannon entropy, and edge density; incorporate deep convolutional feature embeddings alongside interpretable handcrafted descriptors; apply gradient-based continuous optimization over the realism landscape; and validate the framework across additional IQA benchmarks (LIVE, TID2013, SPAQ) to establish broader generalizability.

A. *Acknowledgment*

The authors thank the Department of Computer Science & Engineering, Ashokrao Mane Group of Institutions, Vathar, Maharashtra, India, for providing computational resources and academic support throughout this research.

B. *Funding*

This research received no external funding.

C. *Conflict of Interest*

The authors declare no conflict of interest.

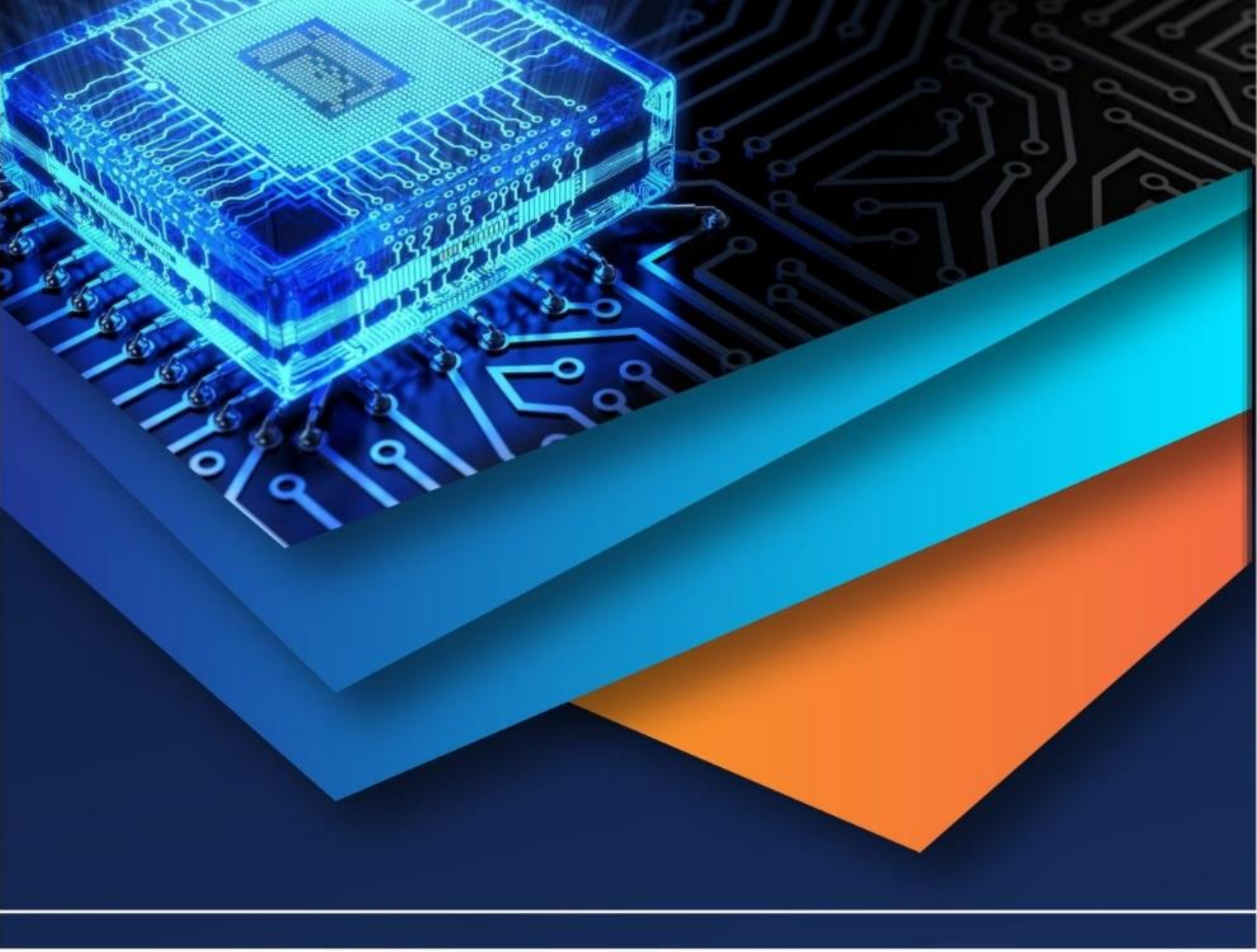
D. *Data Availability*

The KonIQ-10k dataset is publicly available on Kaggle (generalhawking/koniq-10k-dataset). Feature extraction scripts, trained model artifacts, and the Streamlit application code are available from the corresponding author upon reasonable request.

REFERENCES

[1] S. Ma et al., “VQualA 2025 Challenge on Face Image Quality Assessment: Methods and Results,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 3448–3457.
[2] Z. Chen et al., “NTIRE 2025 challenge on real-world face restoration: Methods and results,” in Proceedings of the Computer Vision and Pattern Recognition Conference, 2025, pp. 1536–1547.
[3] K. Misbah et al., “Spatial prediction of soil attributes from PRISMA hyperspectral imagery using wrapper feature selection and ensemble modeling,” PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science, vol. 93, no. 2, pp. 197–215, 2025.
[4] A. Higaki, Y. Kawada, G. Hiasa, T. Yamada, and H. Okayama, “Using a visual Turing test to evaluate the realism of generative adversarial network (GAN)-based synthesized myocardial perfusion images,” Cureus, vol. 14, no. 10, 2022.
[5] W. Alanazi, A. B. Alshaibani, B. A. M. Al-Shaibani, Z. G. Al-Mekhlafi, A. B. Alshaibani, and A. B. Alshaibani, “Machine Learning Models to Identify and Classify Clickbait Headlines Accurately,” Journal of Applied Artificial Intelligence, vol. 6, no. 1, pp. 1–17, 2025.
[6] S. M. A. Al Muhib, R. A. Nayeem, N. Mezi, and N. A. Emon, “A comprehensive image dataset for carambola leaf and fruit disease classification and quality assessment,” Data Brief, vol. 60, p. 111679, 2025.
[7] M. Masih, S. Suleman, M. H. Khan, Z. Sahito, and S. Shahid, “The future classroom: Integrating AI and social media for adaptive learning,” Inverge Journal of Social Sciences, vol. 4, no. 3, pp. 98–111, 2025.
[8] Y. Liu, R. Zhong, Y. Li, C. Yu, J. Zhang, and Z. Kou, “Mapping the evolution of kainate receptor research over five decades: trends, hotspots, and emerging frontiers,” Naunyn. Schmiedebergs. Arch. Pharmacol., vol. 399, no. 2, pp. 2401–2415, 2026.
[9] E. Peli, “Contrast in complex images,” Journal of the optical society of America A, vol. 7, no. 10, pp. 2032–2040, 1990.
[10] A. B. Watson and J. A. Solomon, “Model of visual contrast gain control and pattern masking,” Journal of the optical society of America A, vol. 14, no. 9, pp. 2379–2391, 1997.
[11] R. Sheikh, A. C. Bovik, and G. De Veciana, “An information fidelity criterion for image quality assessment using natural scene statistics,” IEEE Transactions on image processing, vol. 14, no. 12, pp. 2117–2128, 2005.

[12] P. Ye, J. Kumar, L. Kang, and D. Doermann, "Unsupervised feature learning framework for no-reference image quality assessment," in 2012 IEEE conference on computer vision and pattern recognition, 2012, pp. 1098–1105.
[13] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," IEEE Transactions on image processing, vol. 21, no. 12, pp. 4695–4708, 2012.
[14] H. R. Sheikh, M. F. Sabir, A. C. Bovik, and others, "A statistical evaluation of recent full reference image quality assessment algorithms," IEEE Trans. Image Process., vol. 15, no. 11, pp. 3440–3451, 2006.
[15] N. Ponomarenko et al., "Image database TID2013: Peculiarities, results and perspectives," Signal Process. Image Commun., vol. 30, pp. 57–77, 2015.
[16] A. K. Moorthy and A. C. Bovik, "A two-step framework for constructing blind image quality indices," IEEE Signal Process. Lett., vol. 17, no. 5, pp. 513–516, 2010.
[17] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," IEEE transactions on image processing, vol. 23, no. 2, pp. 684–695, 2013.
[18] L. Kang, P. Ye, Y. Li, and D. Doermann, "Convolutional neural networks for no-reference image quality assessment," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, pp. 1733–1740.
[19] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, pp. 586–595.
[20] V. Hosu, H. Lin, T. Sziranyi, and D. Saupe, "KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment," IEEE Transactions on Image Processing, vol. 29, pp. 4041–4056, 2020.
[21] H. A. Pierson and M. S. Gashler, "Deep learning in robotics: a review of recent research," Advanced Robotics, vol. 31, no. 16, pp. 821–835, 2017.
[22] B. Settles, "Active learning literature survey," 2009.
[23] M. Mushtaq, Z. Mehmood, A. A. Butt, M. Y. Shabir, and others, "Causal and Explainable Machine Learning Framework for Heart Disease Prediction using XGBoost and SHAP," Journal of Computing & Biomedical Informatics, vol. 10, no. 01, 2025.
[24] M. C. Granovetter et al., "Functional resilience of the neural visual recognition system post-pediatric occipitotemporal resection," iScience, vol. 27, no. 12, 2024.
[25] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," Mach. Learn., vol. 63, no. 1, pp. 3–42, 2006.
[26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing, vol. 13, no. 4, pp. 600–612, 2004.
[27] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Transactions on image processing, vol. 15, no. 2, pp. 430–444, 2006.
[28] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Blind/referenceless image spatial quality evaluator," in 2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR), 2011, pp. 723–727.
[29] T. O. Aydin, R. Mantiuk, K. Myszkowski, and H.-P. Seidel, "Dynamic range independent image quality assessment," ACM Transactions on Graphics (TOG), vol. 27, no. 3, pp. 1–10, 2008.
[30] C. Guo et al., "Longitudinal evaluation of workflow optimization in radiotherapy: A 4-year retrospective study," J. Appl. Clin. Med. Phys., vol. 26, no. 9, p. e70252, 2025.
[31] T. Ma, S. Pan, J. Guan, Q. Tang, and G. Wang, "Augmented and virtual reality in surgical practice: a review of applications in visualization, training, and future directions," J. Robot. Surg., vol. 20, no. 1, p. 163, 2026.
[32] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, and G.-Z. Yang, "XAI—Explainable artificial intelligence," Sci. Robot., vol. 4, no. 37, p. eaay7120, 2019.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY