ROLE OF VISUAL SALIENCY IN VIDEO QUALITY ASSESSMENT

SYEDA EESHA-TIR-RAZIA
01-244222-013

A thesis submitted in fulfilment of the
requirements for the award of degree of
Master of Science (Electrical Engineering)

Department of Electrical Engineering

BAHRIA UNIVERSITY ISLAMABAD

OCTOBER 2024

MS-14

Approval for Examination

Scholar's Name : **Syeda Eesha-tir-Razia**
Registration Number: **01-244222-013**
Program of Study: **MS in Electrical Engineering**
Thesis Title: **Role of Visual Saliency in Video Quality Assessment**

It is to certify that the above scholar's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for examination. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index **16**% (excluding self-citation), that is within the permissible limit set by the HEC for the MS degree thesis. I have also found the thesis in a format recognized by the BU for the MS thesis.

Principal Supervisor Signature: _____

Date: 10^{th} October 2024

Name: **Dr. Imran Fareed Nizami**

MS-14A

Author's Declaration

I, **Syeda Eesha-tir-Razia** hereby state that my MS thesis titled

"**Role of Visual Saliency in Video Quality Assessment**"

is my own work and has not been submitted previously by me for taking any degree from this university

Bahria University

or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my graduation, the University has the right to withdraw/cancel my MS degree.

Name of Scholar: **Syeda Eesha-tir-Razia**

Date: **10^{th} October 2024**

MS-14B

Plagiarism Undertaking

I, solemnly declare that research work presented in the thesis titled

"**Role of Visual Saliency in Video Quality Assessment**"

is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the university reserves the right to withdraw / revoke my MS degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of scholars are placed who submitted plagiarized thesis.

Scholar/Author's Sign: _____

Name of Scholar: **Syeda Eesha-tir-Razia**

Acknowledgements

Almighty Allah is the Protector, Guider and Provider. Allah Almighty has blessed me with uncountable blessings and given me the strength to complete my research work in due time. Hardwork and trusting His (Allah Almighty) plan is the key to success. As mentioned in the Holy Quran:

"Man can have nothing but what he strives for"

Qur'an (53:39)

I am thankful to my supervisor, **Dr. Imran Fareed Nizami**. for guiding me throughout the research. His advice and knowledge have influenced my academic development and helped me in comprehension of the subject.

Parents can never be thanked enough for their efforts. My beloved parents, **Mr. Syed Mustaneer** and **Mrs. Syedah Sajida Kazmi**, will have my gratitude for life time. They have invested their time, to provide their children with the best education.

I have always received relentless motivation and support from my dear husband, **Syed Umais Gillani**. His belief in my potential has been my constant inspiration to overcome challenges.

Abstract

Human visual system is attracted to unusual or unique objects in a visual scene. In case of computer vision , Visual Saliency is defined as the property that makes certain objects or regions to stand out in an image or scene . Rapid advancement in technology involves video consumption in various domains . Online streaming , virtual reality and video conferencing are some of the glimpse into era of videos . Thus Video Quality Assessment is crucial in these domains . Visual Saliency plays a vital role in video quality assessment by targeting visually important video regions . This helps in development of more robust methods for evaluating video quality , streaming protocols and video compression algorithms. Moreover , it significantly reduces the complexity of evaluation , time and resources. This research work utilizes Global Contrast based visual saliency for video quality assessment . Four methodologies are proposed: Feature Extraction, Feature Selection, Feature Extraction combined with Feature Selection, and Visual Saliency combined with Feature Extraction and Feature Selection. LIVE Video Quality Challenge Database is utilized to evaluate the proposed methodologies . The results of proposed methodologies indicate remarkable performance when compared with state of the art methods . This indicates that visual saliency and feature selection play a significant role to improve the video quality assessment, while efficiently decreasing the computational time.

TABLE OF CONTENTS

LIST OF TABLE

LIST OF FIGURE

ACRONYMS AND ABBREVIATIONS

OG-IQA	Oriented Gradients Image Quality Assessment
GWH-GLBP	Gradient Weighted Histogram of Local Binary Mode on the Gradient Graph
FRIQUEE	Feature Maps Driven No-Reference Image Quality Evaluator Engine
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator index
CQA	Curvelet transform Quality Assessment
VQC	Video Quality Challenge
MSE	Mean Squared Error
PCC	Pearson Correlation Coefficient
KRCC	Kendall Rank Correlation Coefficient
SROCC	Spearman's Rank Correlation Coefficient
LSTM	long-term Temporal Modeling
ConvNet	Convolutional Neural Network
SVR	Support Vector Regression
KNN	K Nearest Neighbor
NR-VQA	No Reference Video Quality Assessment
FR-VQA	Full Reference Video Quality Assessment
RR-VQA	Reduced Reference Video Quality Assessment

CHAPTER 1

INTRODUCTION

1.1 Overview

Objects that are visually appealing instantly grab human attention . Such objects or regions are termed as "salient" . In an image or video , visual saliency is the property of certain elements or regions within a frame that are prominent and is seen immediately . The research work presented in this thesis works towards incorporating visual saliency in order to improve the results of various regression models [3] . This section focuses on different types of VQA , motivation and problem statement which serves as a baseline for the research work.

1.2 Types of Video Quality Assessment

Video Quality Assessment is performed on the basis of Subjective Quality Assessment and Objective Quality Assessment. In Subjective Quality Assessment, the test subjects assign either a MOS (Mean Observer/Opinion Scores) or DMOS (Difference Mean Observer/Opinion Scores) on the basis of perceived video quality. In our dataset, MOS were provided for the videos. For MOS, greater the score, better the quality of video. In case of lower MOS, video quality will be poor. Meanwhile, DMOS is opposite of MOS. Lower the DMOS, better the quality and greater the DMOS, poor the video quality. MOS or DMOS serve as ground truth for the quality assessment of videos. These are often time consuming. Objective Quality Assessment methods aim to automatically predict MOS. Three main types of Objective Video Quality Assessment are discussed below.

1.2.1 Full Reference Video Quality Assessment (FR-VQA)

As the name suggests, a full or whole reference is available in form of original video. This is then compared with the distorted video to predict the quality degradation. It is useful in applications where output video should meet the quality of original video. This involves video compression and transmission, broadcasting, remote diagnostics such as in medical imaging, quality control in production of video content etc.
Figure1.1 shows a block diagram of steps that illustrates working of a FR-VQA approach. Feature based metrics are extracted from both, the reference video and distorted video. The extracted features are then compared by calculating the differences. This step usually involves calculating the differences, such as Mean Squared Errors (MSE). Finally the quality score is determined by aggregating the differences obtained in previous step. Quality scores

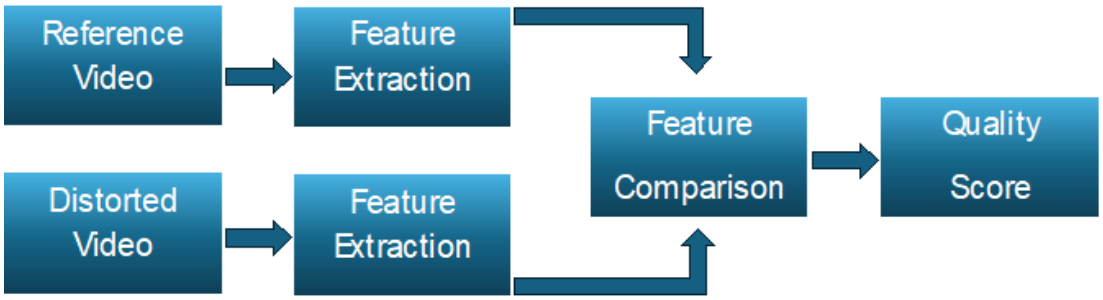


Figure 1.1: Block Diagram of Steps involved in a Full Reference Video Quality Assessment

usually involve either calculating the peak signal-to-noise ratio (PSNR) or more preferably structural similarity index measure (SSIM). SSIM is based on luminance, contrast, and structure similarity between reference and distorted video. SSIM [4] is represented by equation 1.1.

$$SSIM(x,y) = \frac{(2\,\mu_x\mu_y + C_1)(2\sigma_{x,y} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{1.1}$$

x indicates reference video whereas y indicates distorted video. SSIM expression is simply the product of luminance and contrast expressions. Luminance expression involves average of frames μ_x and μ_y. Contrast expression involves standard deviation σ_x and σ_y.

1.2.2 Reduced Reference Video Quality Assessment (RR-VQA)

Reduced reference Video Quality Assessment does not involve full access to the reference video, rather some features or statistics of reference or original video are available. In RR-VQA, the features of reference video and distorted video are combined to predict the quality score.

In [5], authors have presented a Reduced Reference Video Quality Assessment approach. The final Quality score is given by equation 1.2.

$$VQM_{final} = w_{EPSNR} VQM_{EPSNR} + w_{parametric} VQM_{parametric} \tag{1.2}$$

w_{EPSNR} and $w_{parametric}$ are weights for respective video quality metric. Both are given equal weightage of 0.5, since both metrics contribute equally to predict the final video quality metric. $VQM_{parametric}$ is the feature of reference video, which is bit-rate in this case. This is calculated by a piece-wise linear function. Different ranges of bit-rate

are related to the estimated video quality using Mean Opinion Scores. VQM_{EPSNR} is calculated by the equation 1.3 :

$$EPSNR = 10\log_{10}\left(\frac{P^2}{MSE_{edge}}\right) \tag{1.3}$$

where $_P$ denotes the peak pixel value in the edge.

Figure 1.2 shows a block diagram which illustrates the working of a RR-VQA approach. Some information in form of features for a reference video are accessible. Features for distorted video are extracted which are then compared with the features from reference video. At the end, quality score is predicted by aggregating the features that were compared in the previous step.

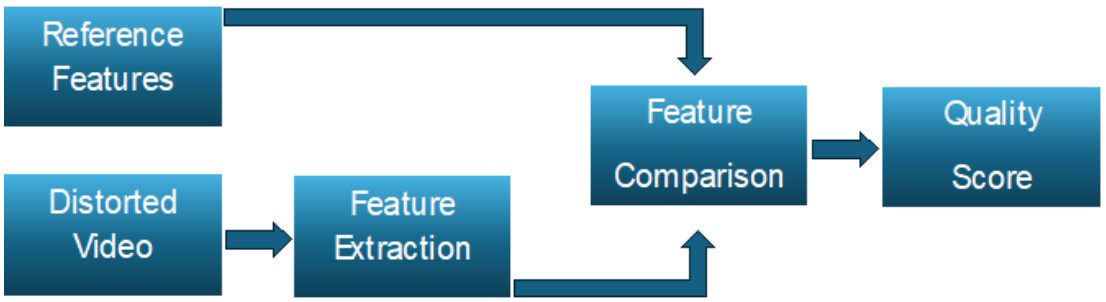


Figure 1.2: Block Diagram of Steps involved in a Reduced Reference Video Quality Assessment

1.2.3 No Reference Video Quality Assessment (NR-VQA)

For a No reference video quality assessment approach, neither a reference video nor a reference feature is available. The distorted video is treated as the first and only information which needs to be targeted. NR-VQA approach is preferred for real world applications such as streaming, user generated content and broadcasting etc.
Figure 1.3 shows a block diagram of steps that occur in a No Reference video quality assessment method. The input is a distorted video for which features are extracted. Features involve both spatial and temporal features for a video. We can use any pre trained model to make predictions using the subjective quality scores from the dataset and then determine the quality score.

1.3 Motivation

Rapid advancement in technology involves video consumption in various domains . Some of the potential motivation behind opting visual saliency in video quality assessment are discussed below:

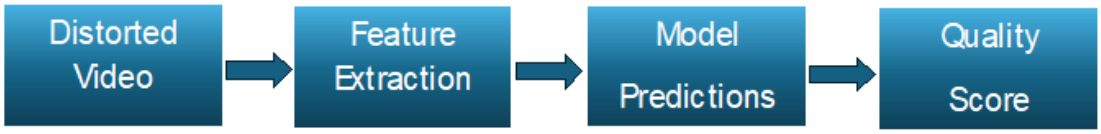


Figure 1.3: Block Diagram of Steps involved in a No Reference Video Quality Assessment

1.3.1 Enhancing Video Quality

Consumption of videos is gradually increasing in today's digital world . Online streaming , virtual reality and video conferencing are some of the glimpse into era of videos . By understanding the role of visual saliency in video quality assessment , more robust methods of evaluating video quality can be developed. This can lead to advancement streaming protocols and video compression algorithms.

1.3.2 Development of New Evaluation Metrics

Existing video quality assessment metrics don't utilize visual saliency for video quality assessment . Visual saliency has been implemented successfully for image quality assessment . But there lies a significant gap in research for implementing visual saliency for video quality assessment problems.

1.3.3 Applications in Computer Vision and Multimedia

Visual saliency is an important concept in computer vision and multimedia analysis . Computer vision applications process large amounts of data from images or videos for a variety of applications . Visual saliency can speed up this process by extracting important information from images or videos for tasks like object recognition , classification , video summarization , content-based video retrieval etc.

1.4 Problem Statement

Existing video quality evaluation strategies frequently neglect to catch the impact of visual saliency, which is a key part of human discernment and consideration . The key components of problem statement is that the focus is on assessing the quality of videos , which includes estimating the apparent constancy , clearness, and in general survey insight . Visual saliency has been implemented for image quality assessment problems . However, visual saliency for Video quality assessment problem needs to be explored further in research . Also , there is a need to identify which factors helps to improve the results of video quality assessment problems , while remaining computationally efficient . The research aims to contribute to the field of video quality assessment by providing insights into the role of visual saliency. For this purpose, four methodologies are proposed

and the results are discussed which imply that visual saliency and feature selection in video quality assessment improves the accuracy and reliability of video quality evaluation.

1.5 Research Objectives

The goal is to assess the viability and benefits of utilizing visual saliency in catching perceptually significant quality variables and how it impacts the results. The main objectives for the research are defined below.

i. Evaluate the performance of various models by performing feature extraction and evaluating the results.

ii. Analyze and compare the affect of feature selection on training time and results.

iii. To analyze the impact of visual saliency on results of models.

1.6 Thesis Organization

The rest of the thesis is organized as follows:

In **Chapter 2**, literature review of different techniques is presented.

In **Chapter 3**, Focuses on proposed methodologies, feature extraction techniques, details of visual saliency and feature selection.

In **Chapter 4**, Experimentation is performed to evaluate the results of proposed methodologies.

In **Chapter 5**, Conclusions and implications are presented.

CHAPTER 2

LITERATURE REVIEW

2.1 Overview

Various research papers have presented methods for image quality assessment problems which forms a road map for video quality assessment problems. Some of the related work and their limitations are discussed in this section.

2.2 Related Work

A FR-VQA approach for Screen Content videos is presented in [6] . Although there has been alot of research in the field of VQA to develop objective metrics for Screen Content Videos but there exist some limitations in terms of generalizability and application of such methologies in diverse video content . Moreover , the research focuses on spatial QA for screen content videos while overlooking important temporal distortions in the videos . Also there is a limited difference between textual , pictorial and computer graphics in screen content videos which overlooks specific quality issues in each region.

In [7], a RR-VQA approach is used for perceptual quality of compressed point clouds . The research was tested on different datasets which introduced over fitting and impacted the generalizability . The proposed Reduced Reference point cloud assessment selectively chooses the important features , but the complexity of feature selection remains a challenge . There is a need for more flexibility in number of features . If not , then those features should be selected which are more important.

A feature extraction method for NR-VQA is presented in [8] . The main limitation is that the multidimensional feature extractor , extracts repeated information which impacts the efficiency of the feature extraction . Also , the input data requires a balance between features which are from subjective and objective description . This negatively impacts the zero-shot feature extraction method, since the amount of information extracted from the video frames is restricted.

Kwong and fellow authors have presented a deep learning based NR-VQA for screen content videos [9] . Training a deep neural network directly on high resolution videos require substantial computational power and memory . These run time requirements limits the proposed model to operate in real time applications.

V-PCC (video-based point cloud compression) Projection Based Blind Point Cloud Quality Assessment for Compression Distortion is a NR-VQA based approach , presented in [10]. Point clouds are three dimensional representation of objects and commonly

used in communication , virtual reality and 3D modelling . Optimizing and fine-tuning the parameters of a dual-stream network can be more challenging than a single-stream networks . In addition , Dual-stream networks may require larger amounts of training data to effectively learn and extract features from both streams . Thus it leads to higher computational complexity and processing time.

Some of the VQA approaches have incorporated visual saliency. Video Multi-method Assessment Fusion (VMAF) presented in [11] , combines multiple features to obtain a score. The perceptual score is based on actual user ratings which aligns with metrics of full reference . According to the authors , the approach utilizes motion quantities . However , instead of relying on motion quantities , temporal features could be extracted from the video . Temporal features provide a comprehensive understanding of motion dynamics and better illustrate how viewers perceive motion in various contexts.

Multi-View Video Quality enhancement Method based on multi-scale fusion convolutional neural network and visual saliency is presented by authors [12] . This is also a FR-VQA approach as indicated by evaluation metrics such as Peak Signal-to-Noise Ratio (PSNR) , Structural Similarity Index (SSIM) , and other objective evaluation indices that compare the processed images to reference images . As discussed before , FR-VQA approaches are not practical for real world problems , since high quality reference content may not be accessible . In addition , the effectiveness of FR-VQA is highly dependent on quality of reference video . If the quality of reference video is poor then the accuracy of the enhancement method will be affected.

Increase efficiency at lower cost (video encoding) using H264 and H265 [13] , Image quality assessment methods are used for video quality assessment problems. Hence there is lack of temporal features which are important for VQA. Train learning based model video features , subjective quality scores [14] . Accuracy of the model when assessing using recent consumer video quality datasets with natural distortions is a challenge.

Double stage visual saliency detection method with pre-attentive features and center bias [15], Still features are only considered in this method.

Visual saliency detection method based on visual center shift [16]. The proposed model cannot provide output with a filled region representing a salient object. An end to end no reference video quality assessment method is presented in [17] which uses a hierarchical spatio-temporal feature representation . It is No Reference (NR) based approach which uses Deep Neural Network (DNN). This method involves two phases i.e. Feature extraction and quality regression which compromises effectiveness of the model . Both phases cannot be optimized simultaneously since there always exists a trade-off. Spatio-temporal(time and change) features occur in two steps. Firstly , spatial features are extracted and then associations are established . Associations are formed only for last layers which hints at the vanishing gradient problem. Ultimately , minute details are lost, which might have been important.

Deep Multimodality Learning based technique for Unmanned aerial vehicle video quality assessment is presented in [18]. It uses ConvNet . The objectives of this research work include, Capture Multimodal features that are extracted from multiple sources such as text , images , audio , etc. These features capture different aspects of data and can provide a more comprehensive representation. The main limitation is the availability of

smaller datasets for video aesthetics . Since deep learning models require larger datasets for better results , hence deep learing cannot be used for this problem . Dataset quality and quantity needs to be improved. Also , Framework should be replaced by neural network to make it end to end.

Learning based Blind video quality assessment models are more versatile in which machine learning regression is employed to map features onto quality score. Challenge is addressed in [19] to utilize Blind video quality assessment for user generated content based video quality assessment. It is based on Two level Transfer Learning for Video Quality Assessment Feature extraction mechanism by using Feature Fusion Video Quality Evaluator while Feature selection using random forest and support vector machine. The limitations are that the Deep learning models which are trained on image datasets are applied for VQA. Additionally, simulated videos are used for VQA instead of real world data. Temporal pooling is an unresolved problem as per the research.

A NR based approach, MD-VQA: Multi-Dimensional Quality Assessment for UGC Live Videos is presented by Zhang [20]. This research work does not incorporate visual saliency. In addition, model incorporates only motion features and semantic distortions and may not account for all potential factors which affect the video quality.

ReLaX-VQA [21], Residual Fragment and Layer Stack Extraction for Enhancing Video Quality Assessment, employed a layer-stacking technique in deep neural networks. The use of deep learning increases complexity and limits real time applicability. Although the model captures salient features through residual fragment extraction, it does not explicitly does not explicitly mention visual saliency as a separate component. The design and feature extraction methods inherently consider saliency by prioritizing areas of the video that are most likely to impact perceived quality. Therefore it may not account for all aspects of visual saliency that influence human perception of video quality.

HVS-5M [22], which stands for a framework that simulates the human visual system (HVS) by incorporating five key characteristics: visual saliency, content-dependency, edge masking, motion perception, and temporal hysteresis. The paper notes that effectively applying visual saliency to video quality assessment remains a challenge. Although the framework includes a visual saliency module, there is still room for improvement in how saliency is utilized to enhance quality predictions.

In [23], an objective quality assessment method is reviewed for image processing and analysis. A detailed analysis of video quality and task based methods used for objective quality assessment of medical data are discussed in the article . Some of the key limitations are that with increase in the depth of network , the performance of neural network based observer decreases . Using loss functions not based on tasks to optimize CNN-based models results in loss of information. Also , Publicly available datasets for medical imaging , particularly in stereoscopic 3D and light field 3D formats , are scarce . Moreover , Computer generated images are used instead of real ones in task based quality assessment.

A Unified Video Quality Assessment Model for User, Professionally, and Occupationally generated content presented by Huang and fellow authors in [24]. As per the proposed model , it is applicable for real time problems. However , various sampling and assessment methods introduce complexity and this can pose challenges to maintain low computation times for different scenarios . Moreover, the model relies on non-uniform

sampling based on frame importance . This approach may not fully capture the nuances of user engagement and perception across different video types. In addition , different visual saliency techniques are used for User , Professionally , and Occupationally-Generated videos . This makes the model more complex and difficult for practical quality assessment . A visual saliency technique that works for one type of video may not be as effective for other types of videos . This also leads to over fitting, where model performs well on training data but but fails to generalize to unseen data.

The main drawback of a FR-VQA is that it requires access to the original or reference video which is not practical for real world problems . It involves comparison of distorted signal with original signal . During the process , If the original signal is lost or not accessible , it will impact the perceived quality . This dependency is the main reason why FR based VQA is not preferable . On the other hand , RR-VQA approach offer its own limitations . Firstly , selection of most effective features from a large number of features is crucial in order to save computational resources . Also, the RR metrics or accuracy rely on those extracted features which predict the quality . This is both complex and challenging . To overcome the challenges faced by FR-VQA and RR-VQA methods , NR-VQA approach is preferred . NR-VQA approach does not require large number of features , rather only effective features serve the purpose . This offers flexibility and practicality where computational resources are limited . Visual saliency improves the accuracy of video assessment problems by aligning them with human perception of video quality . Saliency is incorporated into VQA to guarantee that the assessment is more in line with subjective human quality evaluations . This alignment is essential in order to create models that can accurately forecast how humans perceive a given video quality . By targeting areas of video that are likely to catch human attention , computational load can be reduced. Saliency based VQA models can avoid wasting resources on parts of the video that are not important to perceived quality . This makes the assessment process more efficient and can improve real-time processing capabilities.

To overcome the limitations addressed in related work, we have presented a NR-VQA based approach to evaluate the role of visual saliency in video quality assessment . We have also incorporated feature selection in our work to demonstrate how it affects the video quality assessment results , when combined with visual saliency. While visual saliency concentrates the assessment on perceptually significant regions , feature selection guarantees that only the most pertinent features are employed. This combination optimizes accuracy and efficiency. The VQA is more scalable and economical when computational efforts are concentrated on salient regions and pertinent features , hence minimizing total resource consumption . The model is more resilient in a variety of video settings when features are chosen carefully to lower the chance of overfitting and visual saliency adjusts to different kinds of information.

CHAPTER 3

METHODOLOGY

3.1 Overview

In this section, Visual saliency technique, feature extraction techniques, details about feature selection and proposed methodolgies are discussed.

3.2 Global Contrast Based Visual Saliency (GC)

There are various visual saliency algorithms [25], [26], [27] . Global contrast-based visual saliency is an effective and biologically inspired technique for identifying key regions in images . By emphasizing global differences in features such as color , intensity , and texture , it successfully replicates the human visual system's ability to detect prominent objects [28] . This approach is extensively applied in numerous fields , including object detection, image segmentation , robotics , and human-computer interaction , highlighting its adaptability and efficacy in visual analysis tasks.

In global contrast-based saliency , region or pixel is compared to the whole image which is termed as global contrast. This measures the degree of deviation of the localized region with reference to overall statistics of the image (such as color , intensity , texture etc). If the color, brightness , or roughness of a pixel is quite different from what is common with other pixels in the given image, then it is said to be salient . It is normally characterized by the creation of a saliency map that shows a quantitative value at every pixel , depicting the saliency degree . Lighter areas have higher values which indicate salient region.

3.2.1 Steps in Global-Contrast Based Visual Saliency

Figure 3.1 illustrates the working of GC based Visual Saliency algorithm. Given below are details of each step.

3.2.1.1 Pre-processing with Color Space

Input frame is generally pre-processed. For example , resizing and converting the frame to the desired color space and so on. It is worth to mention that the color space for the frames influences the calculation of saliency . Such setups as LAB that withdraw tint from brightness can be of great assistance in emphasizing color differences.

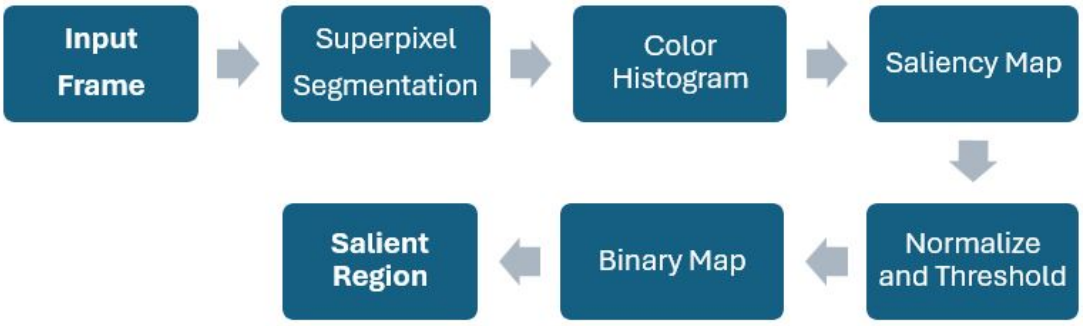


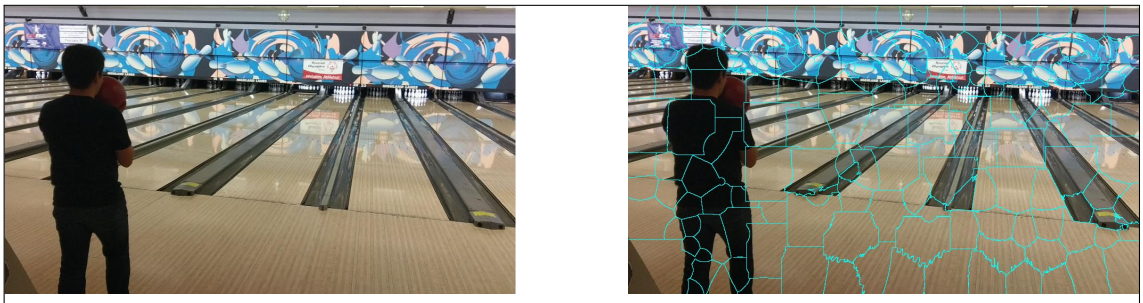
Figure 3.1: Steps In GC based Visual Saliency

3.2.1.2 Image Segmentation

The video frame is divided into regions called super pixels [29], [30]. For each region, Basic features of the input frame are extracted. These include: color, tone, intensity and texture.

- Color: Certain colours stand out in a frame. For example, red in comparison with green. These variations can provide a color contrast for a region.
- Texture: Certain trends or fluctuations in a frame reveal geometric properties or patterns.
- Intensity: Changes in brightness or fire capacity.

Super-pixel segmentation algorithm has certain parameters of tuning which actually decides the size of the super-pixels along with their compactness . These parameters are used in the saliency map and can influence the quality of the map . The parameters can only be selected with considerable care and this in most cases involves testing . In this case , super pixels are defined as 200.



(a) Input Frame (b) Super-pixel Segmentation

Figure 3.2: Segmentation of an Image in Super-pixels

3.2.1.3 Histogram Computation

A color histogram is a graphical display for the intensity of pixel colors , within each region of a frame . In terms of the LAB color , histogram reflects the distribution of pixels in the given L, A, and B ranges.

- L indicates lightness, with values varying between 0 and 100.
- A: An intermediate color-opponent dimension that varies from green to red.
- B: Opacity dimension, which is the color-opponent dimension that ranges from blue to yellow.

Figure 3.3 shows histograms of first two regions of a frame. Histogram for L , A and B are computed for region 1, in which bins on the horizontal axis of the histogram indicate pixel ranges. Each of the color channel (L,A,B) is divided into 16 intervals , where each interval is taken to correspond to some range of values in the corresponding color channel (L,A,B). The y-axis indicates the frequency of occurrence or the count of pixels in the specified bin . This count is sometimes recapitulated to show proportions or probabilities (sum to 1).

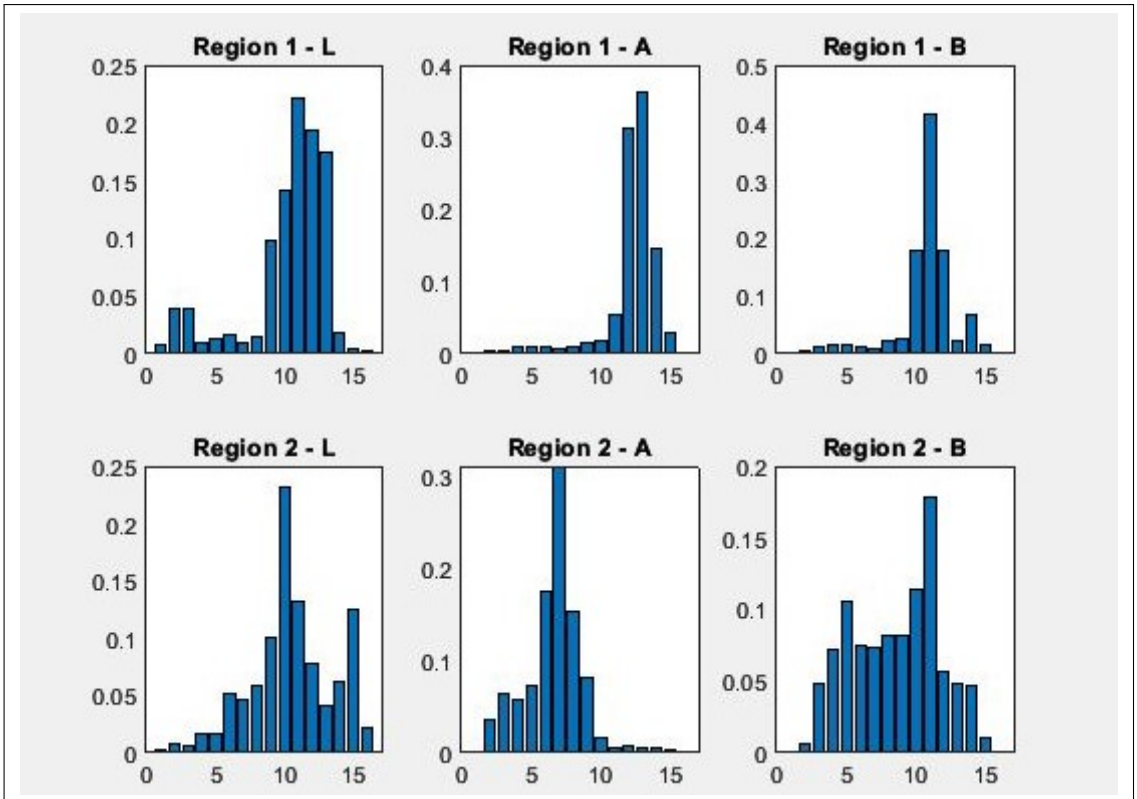


Figure 3.3: LAB Color Histograms For Two Regions of a Frame

3.2.1.4 Computing Saliency Map

Sum of squared differences between histogram and the histograms of all other superpixels is used to obtain Color distance metric. Formulas for saliency values and color distance metric [28] are given below. The value of the saliency map is the sum of these distances, which means that the saliency of region is greater if it has a higher degree of difference from other regions [31]. Each region is assigned a saliency value which is then mapped on the image to obtain saliency map. The saliency value for a region r_k with respect to other regions r_i is given by equation 3.1,

$$S(r_k) = \sum_{r_k \neq r_i} w(r_i) D_r(r_k, r_i) \tag{3.1}$$

where:

$w(r_i) =$ Weight of region r_i

$D_r =$ Color distance metric between the two regions

$$D_r(r_1, r_2) = \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} f(c_{1,i}) f(c_{2,j}) D(c_{1,i}, c_{2,j}) \tag{3.2}$$

where:

$k = 1,2$

$f(ck_{,i}) =$ Probability of occurrence of the i-th color in region ck

$n_k =$ All colors in the region r_k

$D_r =$ color distance metric between region $r_1$and region r_2

$n_1, n_2 =$ the number of intervals or bins in the color histogram of region $r_1$and r_2

$D(c_{k,i}) =$ Distance between two colors $(c_{1,i})$and $(c_{2,j})$

Saliency maps for a random frame of three videos is shown in Figure 3.4. (a), (b) and (c), are random frames (out of 300) for three different videos, whereas, (d), (e) and (f) show their respective Saliency Maps.

3.2.1.5 Binary Map

To obtain final frame representing the salient region, saliency map is converted into a binary map . A threshold is defined and the super pixels that were previously mapped on the frame are compared with the threshold value . The superpixels which have values greater than the threshold, will be assigned the boolean value of true 1 and all other pixels will get the boolean value of false 0 . Last of all, This step superimposes the binarized 'map' on top of the original frame to highlight the salient region.

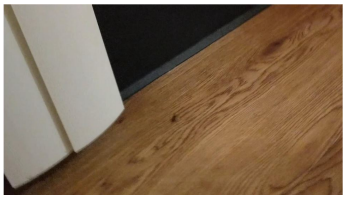
3.3 Feature Extraction Techniques

3.3.1 OG-IQA

OG-IQA is a feature descriptor technique used in image pre-processing. A publicly available OG-IQA [32] has been used for feature extraction . The OGIQA technique analyzes an object's shape and appearance by looking at the distribution of edge orientations within it. The process , called OG-IQA , specifically comprises figuring out the gradient magnitude. On the basis of position of pixels, the image is then divided into smaller cells in order to conduct additional analysis . There are six features total that are taken

(a) Frame of Video 1

(d) Saliency Map 1



(b) Frame of Video 2

(e) Saliency Map 2

(c) Frame of Video 3

(f) Saliency Map 3

Figure 3.4: Frames of three videos and their saliency maps

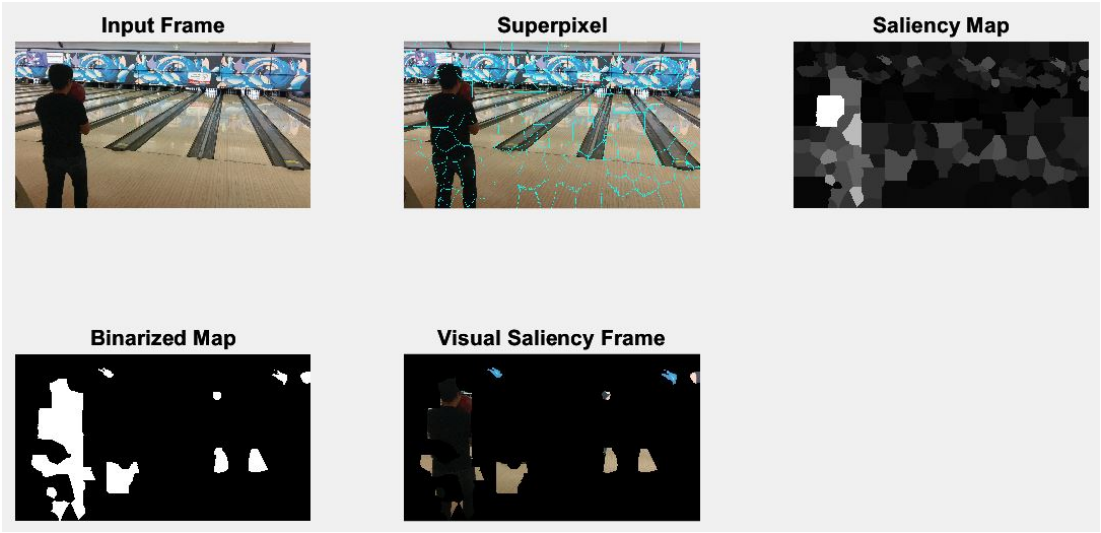


Figure 3.5: Steps to Obtain a Visual Saliency Frame

from each frame . With the help of this innovative approach , the Non-Reference Image Quality Assessment (NR IQA) can now investigate image gradients in more detail than just gradient magnitude . It makes use of the image gradient's orientation , moving the emphasis from magnitude to orientation for NR IQA objectives [33].

The proposed feature extraction technique calculates the gradient orientation for the entire image . The image can either be gray or colored image . This means that not only it calculates the magnitude but also the direction or angle where intensity changes . The gradient information is particularly significant at edges , where there are abrupt changes in intensity , leading to more pronounced gradient values.

Three parameters are calculated while performing feature extraction. Gradient magnitude (GM), relative gradient magnitude (RM), and orientation of relative gradient (RO). The expressions for the equations 3.3 to 3.6 are available in [34]. The expression for calculating gradient magnitude (GM) of the image, at the pixel location (u, v) and orientation of gradient is given by equation 3.3 and equation 3.4 respectively.

$$|I(u,v)| = \sqrt{I_x^2(a,b) + I_y^2(a,b)} \tag{3.3}$$

where:

$I_x^2(a,b)$ = Directional derivative values of horizontal axes in the image
$I_y^2(a,b)$ = Directional derivative values of vertical axes in the image

Whereas the gradient orientation is given by:

$$|\angle(u,v)| = arctan\left(\frac{I_y(a,b)}{I_x(a,b)}\right) \tag{3.4}$$

For a pixel at (u, v) coordinates , the average local orientation is computed by averaging the orientations of all pixels in a defined neighborhood. As a result , every pixel within that neighborhood will have the same local average orientation value assigned to it . This normalization helps to highlight the differences in orientation relative to the surrounding pixels , making it easier to detect distortions or significant features in the image . Relative Gradient Orientation (RO) measures how a specific pixel's gradient orientation deviates from local average . It is useful for detecting anomalies or significant features that stand out from the surrounding context. Expression for RO is given by equation 3.5.

$$\angle(u,v)_{RO} = \angle(u,v) - \angle(u,v)_{AVE} \tag{3.5}$$

Where,

$\angle(u,v)_{AVE}$ = Local average orientation

Relative Gradient Magnitude (RM) is calculated by the expression given in equation 3.6. The resulting feature vector is combination of gradient magnitude (GM), relative gradient magnitude (RM), and relative gradient orientation (RO).

$$|I(u,v)|_{RM} = \sqrt{(I_x(u,v)+I_x(u,v)_{AVE})^2+(I_y(a,b)+I_y(a,b)_{AVE})^2} \tag{3.6}$$

It is possible to develop a robust VQA method that leverages both spatial and temporal information . This approach can enhance the ability to assess video quality in a way that reflects human perception , making it suitable for various applications. For each detected object in a frame , we can calculate the gradient features using the same formulas as in the original research . This will provide a feature set for each object in the frame.

3.3.2 GWH-GLBP-BIQA

A no-reference (NR) Image Quality Assessment (IQA) technique aimed at predicting the visual quality of multiply-distorted images by focusing on structural degradation [35]. This approach involves extracting a novel structural feature known as the gradient-weighted histogram of local binary pattern (LBP) calculated on the gradient map (GWH-GLBP) . This feature proves effective in capturing the intricate degradation patterns induced by multiple distortions . Through extensive experimentation on two publicly available databases containing multiply-distorted images , proposed GWH-GLBP metric has demonstrated superior performance compared to existing full-reference and NR IQA methods , exhibiting high consistency with human subjective ratings . Total number of features extracted per frame are fifty.

The input image is treated as a grayscale image for the purpose of feature extraction . The distorted image is filtered using the Prewitt operator to extract the gradient magnitudes of each pixel in the input image . The gradient magnitudes are calculated by convolving the image with Prewitt filters. For each pixel in the image , one gradient magnitude is computed, resulting in a gradient magnitude map that has the same dimensions as the input image . Local Binary Pattern (LBP) operator is applied to the gradient magnitude map to encode the image's primitive micro structures , such as edges , lines , and spots . This results in the Gradient Local Binary Pattern (GLBP) . The proposed feature extraction method accumulates the gradient magnitudes of pixels with the same GLBP pattern , instead of using a traditional frequency histogram to describe the global structural information . This creates a gradient-weighted GLBP histogram , which emphasizes regions with high contrast changes and combines structural and contrast information . The equations are available in [35]. The equation 3.7 and 3.8 is used to calculate gradient-weighted GLBP . The bins of histogram represent different GLBP patterns and they accumulate gradient magnitudes of the pixels corresponding to those patterns.

$$h_{glbp}(k) = \sum_{i=1}^{N} \omega_i f(\mathrm{GLBP}_{P,R}(i),k) \tag{3.7}$$

Where,

N = Denotes the number of image pixels

k = Represents the possible GLBP patterns

ω_i = Weight assigned to gradient magnitude, calculated from the image

$$f(x,y) = \begin{cases} 1, & x = y \\ 0, & \text{otherwise} \end{cases} \tag{3.8}$$

The above function is used to determine whether the GLBP code at a specific pixel, denoted as x matches a particular GLBP pattern, which is denoted as y. For each pixel in the image, a neighborhood is defined . The intensity values of the surrounding pixels are compared to the intensity value of the central pixel . If the neighbor's intensity is greater than or equal to the central pixel's intensity, it is assigned a value of 1. If the neighbor's intensity is less than the central pixel's intensity , it is assigned a value of 0. By encoding the local binary patterns for all pixels in the image, LBP provides a compact representation of the image's texture and structural features , making it useful for image quality assessment . The LBP codes can then be aggregated to form histograms that summarize the texture information across the entire image or specific regions.

Similar to the IQA approach , a machine learning model can be trained to predict video quality . We can use a dataset of videos with known quality ratings (either subjective ratings from viewers or objective metrics) . Since video consists of a sequence of frames , it is important to consider the temporal aspect. We can aggregating the features from multiple frames to create a comprehensive representation of the video . This may involve averaging the histograms or using statistical measures (e.g., mean, variance) across frames.

3.3.3 FRIQUEE

Images frequently show complex mixtures of different distortions in real-world situations . An innovative method known as the "bag of feature maps" , which refrains from assuming any certain kind or types of distortion to be present in an image . Rather , its emphasis lies on identifying patterns or non-patterns in the statistical characteristics of actual images [36]. A publicly available technique [37] has been used for feature extraction . A regressor is used to forecast image quality by utilizing a sizable dataset of realistically distorted images , associated human ratings , and computed feature bags.

Different color spaces are used to transform the input image . This covers HSI , CIELAB , and RGB . This helps to examine the image in detail , in terms of brightness and chrominance (color). For each image , feature maps are created in various color spaces and transform domains . These feature maps are intended to record different aspects of the perceptual quality and statistics of the image . The feature maps are normalized to reduce the statistical dependency . For each color space, luminance is calculated for all pixels . Normalization is applied to compute the normalized luminance coefficients (NLC)

. This helps in reducing the impact of extreme values and making the data more consistent . Equation 3.9 is used to calculate normalized luminance coefficients (NLC).

$$NLC(i,j) = \frac{L(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{3.9}$$

Where,

$L(i,j)$ = Luminance value at pixel coordinates (i, j)
$\mu(i,j)$ = Local mean luminance around the pixel (i, j)
$\sigma(i,j)$ = Local standard deviation of luminance around the pixel (i, j)

From the NLC map , various statistical features are extracted . These features may include shape parameters (e.g., the shape parameter of a generalized Gaussian distribution) , Variance and standard deviation , Kurtosis and skewness, which provide insights into the distribution of the NLC values. Features are extracted for the original image as well as a decreased resolution of the image (downsampled by a factor of 2) . The utilization of a multiscale technique facilitates the capture of distortion effects at various scales . The feature extraction includes Luminance features (e.g. , normalized luminance coefficients, neighboring paired products, sigma maps), Chroma features (e.g., chroma maps, neighboring pair product maps) , LMS color space features and hue/saturation component features . The features extracted from various color spaces are combined into a large feature set. The goal of this feature extraction approach is to extract a wide range of perceptually meaningful elements that are useful in predicting the quality of images impacted by intricate, real-world aberrations. 560 features are extracted from various feature maps designed for image quality assessment . These features are derived from different color spaces and statistical models , including luminance , chroma , and LMS color spaces , as well as from various perceptually relevant statistical image features . Videos often contain complex and authentic distortions due to compression, transmission errors, and environmental factors. A large feature set enables the assessment model to better differentiate between high-quality and distorted frames, improving its robustness against various types of distortions . As we will see in experimentation, this feature extraction technique performs exceptionally good for VQA as compared to other feature extraction techniques.

3.3.4 Curvelet QA

A two-step approach that begins by detecting any distortions in an image and then evaluates the quality of each distortion separately . In both phases , perceptual characteristics are retrieved and used as training data . Especially, characteristics taken from the curvelet domain show a good correlation with the quality of natural images in all categories of distortion. The computed image curvelet representation yields several statistical properties , such as the coordinates of the log-histogram maxima of the curvelet coefficient values and the energy 9 distributions of the orientation and scale inside the curvelet domain . Results imply that these characteristics successfully capture the existence

and degree of visual distortion [38] . Twelve features are extracted per frame in this technique.

The input image is divided into blocks of size 256 x 256 . This size is large enough to capture significant features of the image while still being small enough to allow for efficient processing and analysis . For each 256 x 256 block , the curvelet transform is applied. The curvelet transform can be thought of as a multi dimensional extension of the wavelet transform , designed to capture directional information and singularities more effectively . This means that for each block, the transform generates coefficients that represent the image at different levels of detail . Eqaution 3.10 shows formula to calculate curvelet transform of a 2-D function [38]. The scale, orientation and position determines size , direction and location of the curvelet respectively.

$$\theta(j,l,k) = \sum_{(t_1,t_2)} f(t_1,t_2) \cdot \phi_{j,l,k}(t_1,t_2) \tag{3.10}$$

Where:

$$\begin{aligned} \theta(j,l,k) &= \text{Curvelet coefficients at scale } j, \text{ orientation } l, \text{and position } k \\ f(t_1,t_2) &= \text{Input image} \\ \phi_{j,l,k}(t_1,t_2) &= \text{Curvelet function corresponding to scale, orientation, and position} \end{aligned}$$

At the coarsest scale , the transform captures the overall structure of the block . As the scale increases , the transform captures finer details , such as edges and textures. The curvelet transform is particularly adept at representing curves and edges due to its directional sensitivity. The result of the curvelet transform is a set of curvelet coefficients for each block. These coefficients are organized in a way that reflects both the scale and orientation of the features in the image . Each coefficient corresponds to a specific curvelet, which is parameterized by its position , scale, and orientation. The coefficients are larger when the curvelet aligns well with the features in the image (e.g., edges) , and smaller when there is misalignment . One of the key advantages of the curvelet transform is that it provides a sparse representation of the image . This means that only a small number of coefficients are needed to accurately represent significant features of the image, particularly those with singularities . After obtaining the curvelet coefficients , various statistical features can be extracted from them . These features may include empirical probability distribution function (PDF) of the logarithm of the coefficients , fitting parameters such as amplitude , mean , standard deviation and orientation energy distributions that capture the directional information of the coefficients . The extracted features from all blocks are then averaged or combined to create a final feature vector that represents the quality of the entire image.

Similar to the block-based approach used in images, the video frames can be divided into blocks . Features can be computed for each block in every frame , allowing for localized quality assessment . This can help in identifying specific regions in the video that may suffer from quality degradation.

3.3.5 BRISQUE

BRISQUE [39] is a spatially-based, blind/no-reference (NR) image quality assessment (IQA) model that is based on natural scene statistics . In contrast to techniques that calculate distortion-specific features such as blocking, blur, or ringing , BRISQUE focuses on assessing the potential loss of "naturalness" in the image owing to distortions by using scene statistics of locally normalized brightness coefficients . This method offers a thorough assessment of quality. Based on a spatial natural scene statistic model , the model uses features that are obtained from the empirical distribution of locally normalized luminances and products of locally normalized luminances . Unlike other NR IQA methods , BRISQUE notably does not require transformation to another coordinate frame . BRISQUE is accessible to the general public [40]. Thirty six features are extracted in this technique per frame.

Feature extraction process involves computing Mean Subtracted Contrast Normalized (MSCN) coefficients , fitting statistical models to these coefficients , and extracting features from both the coefficients and their pairwise products across multiple scales. locally normalized luminances from the input image are computed by using local mean subtraction and divisive normalization . This is done to remove local mean displacements and normalize the local variance of the log contrast . The normalized luminance values are referred to as Mean Subtracted Contrast Normalized (MSCN) coefficients . This transformation helps in modeling the contrast gain masking process in early human vision . MSCN [39] can be calculated by equation 3.11 . This formula is used to normalize the luminance values of the image, allowing for the extraction of features that capture the statistical properties of the image in the presence of distortions.

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{3.11}$$

$$\begin{aligned} \hat{I}(i,j) &= \text{MSCN coefficient at pixel location (i, j)} \\ \mu(i,j)) &= \text{local mean of the image around pixel (i, j)} \\ \sigma(i,j) &= \text{local standard deviation of the image around pixel (i, j)} \\ C &= \text{small constant added to avoid division by zero} \end{aligned}$$

The research quantifies these changes to predict the type of distortion and the perceptual quality of the image . The empirical distributions of the MSCN coefficients are fitted to a generalized Gaussian distribution (GGD) to capture the statistical properties of both pristine and distorted images . Two parameters (shape and scale) are estimated from this fit , forming the first set of features. Additional features are extracted by computing pairwise products of the MSCN coefficients along four orientations (horizontal , vertical , main-diagonal , and secondary-diagonal) . These products are also fitted to an asymmetric generalized Gaussian distribution (AGGD), yielding further parameters that contribute to the feature set . Features are extracted at the original image scale and at reduced resolution

scale (low pass filtered and downsampled) . A total of 36 features (18 at each scale) are used to identify distortions and perform distortion-specific quality assessment.

The methodology can be adapted to various video formats and compression standards , making it versatile for different applications in video processing and transmission . Once the BRISQUE features are extracted for each frame , the resulting quality scores can be aggregated to produce an overall video quality score . This can be done using various methods , such as averaging the scores of all frames or applying a weighted approach that considers the importance of certain frames.

All of the feature extraction techniques mentioned above , have been used in this research work. This helps to evaluate the best feature extraction technique for video quality assessment.

3.4 Feature Selection

The process of taking a meaningful subset of features (attributes) out of the original set of features , in a dataset is called feature selection. It seeks to remove superfluous , redundant , or noisy data while identifying the most important characteristics that enhance a model's predicted accuracy [41]. The performance of data mining algorithms , computational efficiency , and the interpretability of the output can all be improved by this feature reduction . In many data mining tasks, including regression, association rules , clustering , and classification , feature selection is a crucial stage in the preparation of the data.

3.4.1 Waikato Environment for Knowledge Analysis (WEKA)

A collection of open source machine learning and data analysis tools known as (Weka) [42] is available under the General Public License . Weka has a number of visualization tools, algorithms , and graphical user interfaces for functions related to data analysis and predictive modeling . Data preprocessing , clustering , classification , regression , visualization , and feature selection are among the common data mining activities that Weka can do . Various research works have used WEKA for feature selection [43], [44]. It is anticipated that files submitted to Weka will be named with the.arff extension and formatted using the Attribute-Relational File Format . Weka's methods are based on the supposition that the data is provided in the form of a single flat file or relation , with each data item having a fixed set of attributes (usually nominal or numeric properties). Interface of the WEKA tool is shown in Figure 3.6.

A comma-separated values (CSV) file is uploaded in which features and labels are available. The last column contains MOS (Mean opinion Scores) as labels. In WEKA, Under select attributes option , we can choose an evaluator and a search method. In our case , CfsSubset was used as evaluator with genetic search algorithm as search method. According to the details of CfsSubset evaluator mentioned in [45] , it evaluates the effectiveness of feature subsets by taking into account two primary aspects , the degree of inter-correlation among the features and the value of each feature in predicting the class label , which is MOS in our case . Good feature subsets , according to the underlying hypothesis , it should include features that are uncorrelated with one another and highly correlated with the class.

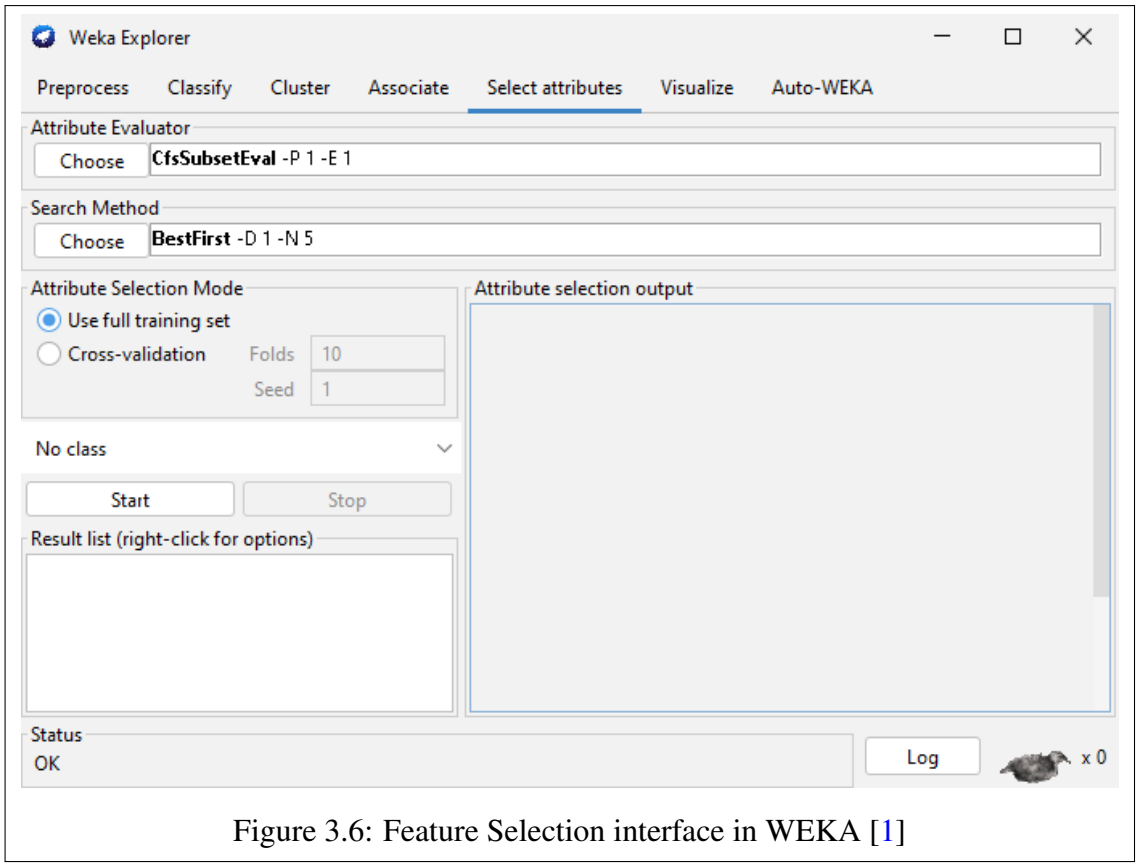


Figure 3.6: Feature Selection interface in WEKA [1]

3.5 Proposed Methodology

On the basis of details mentioned in the previous sections, we propose four different methodologies to demonstrate how they affect the results of VQA.

3.5.1 Proposed Methodology 1- Feature Extraction

Figure 3.7 shows block diagram of the proposed methodology. We extracted 300 frames per video , for 585 videos. For each frame, features are extracted using different feature extraction techniques , which are already discussed in Section 3.3. For FRIQUEE , 560 features are obtained , OGIQA gives 6 features, BRISQUE extracts 36 features, Curvelet extracts 12, and GWH-GLBP extracts 50 features per frame . For each feature extraction technique , extracted features for all videos were saved as csv files . These csv files were evaluated on different regression models , Support Vector Regression , K Nearest Neighbor, Random Forest , Decision tree , Linear regression and ConvNet.

3.5.2 Proposed Methodology 2- Feature Selection

Figure 3.8 shows block diagram of the proposed methodology . The initial step of feature extraction is same as previous methodology. We extracted 300 frames per video, for 585 videos. For each frame , features are extracted using different feature extraction techniques , which are already discussed in Section 3.3. For FRIQUEE , 560 features are obtained , OGIQA gives 6 features , BRISQUE extracts 36 features, Curvelet extracts 12,

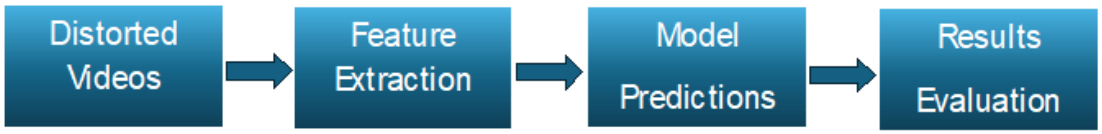


Figure 3.7: Block diagram which demonstrates steps of Proposed Methodology 1- Feature Extraction

and GWH-GLBP extracts 50 features per frame . For each feature extraction technique , extracted features for all videos were saved as csv files.

The difference in this methodology is that we combine all the extracted feature csv files into a single csv file . For this merged csv file , we perform feature selection using WEKA, details of which are covered in Subsection 3.4.1. The important feature columns are kept while redundant feature columns are dropped . This updated feature file is evaluated on different regression models for evaluation . Same regression models are used which are considered in previous methodology . It includes Support Vector Regression, K Nearest Neighbor , Random Forest , Decision tree, Linear regression and ConvNet.

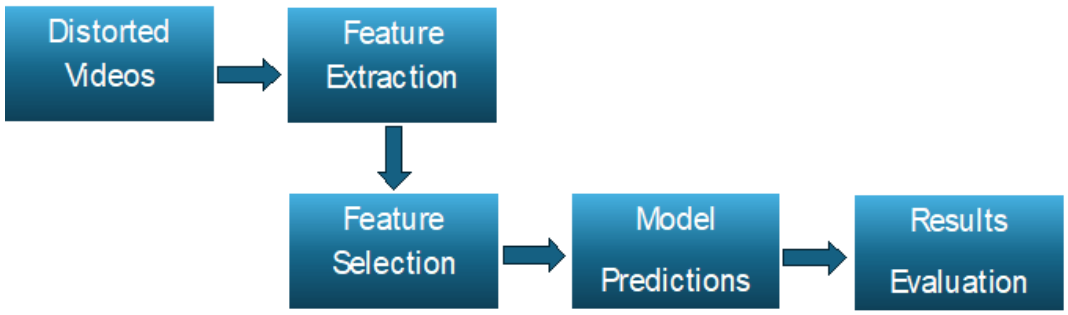


Figure 3.8: Block diagram which demonstrates steps of Proposed Methodology 2- Feature Selection

3.5.3 Proposed Methodology 3- Visual Saliency

Figure 3.9 shows block diagram of the proposed methodology.

In this methodology , we employ Global Contrast based visual saliency. The details about GC based visual saliency are covered in Section 3.2. Visual Saliency is applied to video frames which highlights the salient regions inside frames. The visual saliency frames are converted into binary maps. The white regions indicate salient regions whereas the black regions are considered as non salient.

We perform feature extraction for these visually salient frames . Rest of the process is same as previous methodology. Feature extraction is performed for all frames . We extracted 300 frames per video , for 585 videos . For each frame, features are extracted using different feature extraction techniques, which are already discussed in Section 3.3.

For FRIQUEE , 560 features are obtained , OGIQA gives 6 features , BRISQUE extracts 36 features, Curvelet extracts 12 , and GWH-GLBP extracts 50 features per frame. For each feature extraction technique, extracted features for all frames were saved as csv files . These files were evaluated on different regression models for evaluation . Same regression models are used which are considered in previous methodologies . It includes Support Vector Regression, K Nearest Neighbor, Random Forest, Decision tree, Linear regression and ConvNet.

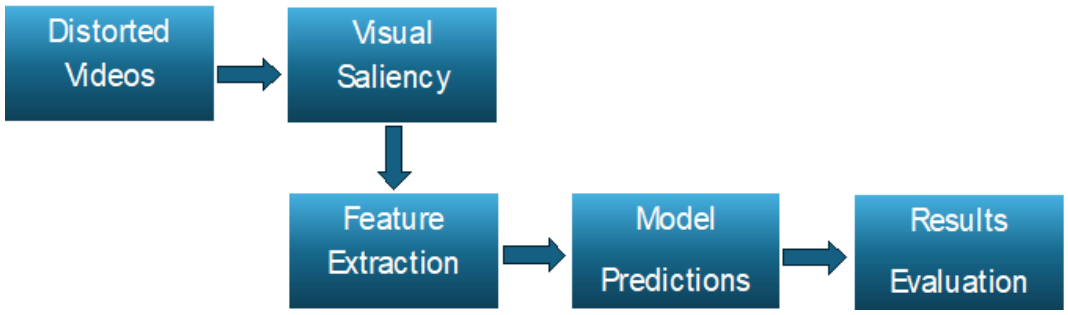


Figure 3.9: Block diagram which demonstrates steps of Proposed Methodology 3- Visual Saliency

3.5.4 Proposed Methodology 4- Visual Saliency, Feature Extraction combined with Feature Selection

This methodology combines all previous methodolgies . Figure 3.10 shows block diagram of the proposed methodology. Visual Saliency is applied to video frames which highlights the salient regions inside frames . The visual saliency frames are converted into binary maps. The white regions indicate salient regions whereas the black regions are considered as non salient . We perform feature extraction for these visually salient frames . For each frame , features are extracted using different feature extraction techniques, which are already discussed in Section 3.3. For FRIQUEE , 560 features are obtained , OGIQA gives 6 features, BRISQUE extracts 36 features, Curvelet extracts 12, and GWH-GLBP extracts 50 features per frame. For each feature extraction technique , extracted features for all frames were saved as csv files. In feature selection step , We combine all the extracted feature csv files into a single csv file. For this merged csv file , we perform feature selection using WEKA. The important feature columns are kept while redundant feature columns are dropped . This updated feature file is evaluated on different regression models for evaluation . Same regression models are used which are considered in previous methodology. It includes Support Vector Regression, K Nearest Neighbor, Random Forest, Decision tree, Linear regression and ConvNet.

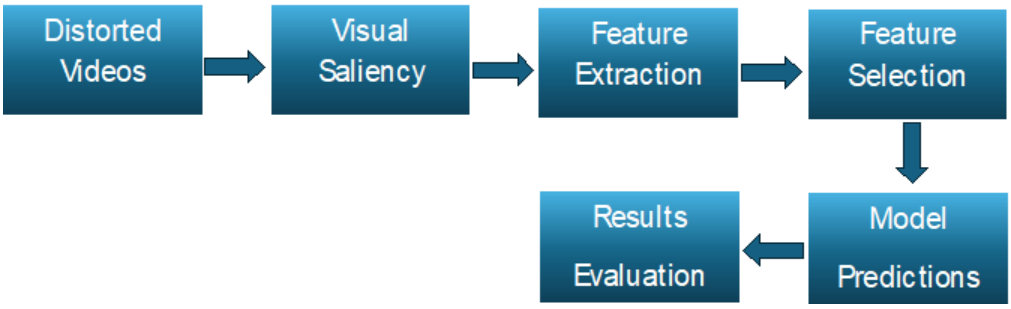


Figure 3.10: Block diagram which demonstrates steps of Proposed Methodology 4- Visual Saliency, Feature Extraction combined with Feature Selection

CHAPTER 4

EXPERIMENTATION AND RESULTS

4.1 Overview

In this section, the results of regression models in terms of various evaluation metrics are discussed. The initial experimentation is performed without incorporating visual saliency. The frames and feature extraction for videos is performed in MATLAB. The feature extraction is performed for each frame using OGIQA, BRISQUE, FRIQUEE, CURVELET and GWH-GLBP. The features extracted by these techniques are then evaluated on different regression models, which is performed in Spyder integrated development environment.

4.2 Dataset

A comprehensive database for assessing video quality on a large scale , comprising 585 videos with distinct content, recorded using 101 diverse devices (from 43 different models) by 80 individual users , featuring a wide spectrum of complex and authentic distortions [46], [47]. A screenshot of dataset is shown in Figure 4.1.

Figure 4.1: LIVE Video Quality Challenge Database [2]

It contains 585 videos of unique content captured by a large number of users , which ensures a wide variety of video types and scenarios . This diversity helps VQA models to generalize better across different video contents . It includes over 205,000 opinion scores collected from 4,776 unique participants through crowdsourcing . This large volume of subjective ratings provides a robust basis for evaluating and training VQA models . The dataset also contains mean observer scores (MOS) for all videos . Also , videos in the

dataset are impaired by complex , authentic distortions that occur in real-world scenarios , including poor exposures , motion blurs , color representation issues , and more . This variety allows researchers to test and improve VQA algorithms against a comprehensive set of challenges . making it more representative of actual video quality challenges compared to traditional datasets that often use synthetic distortions . By providing a large-scale, high-quality dataset, it facilitates further research in the field of VQA , encouraging the development of more accurate and reliable no-reference video quality assessment models.

4.3 Evaluation Parameters

Regression models are evaluated on different correlation parameters. Correlation coefficients are used in the context of video or image quality assessment to determine the link between subjective evaluations (Mean Observes Scores) and objective quality measures (NR-VQA approach in our case). This can help determine whether features have a greater impact on perceived quality. The details of KRCC, SROCC, MSE and PCC are discussed below.

4.3.1 MSE

A metric called mean squared error (MSE) is used to calculate the average of the squares of the errors , or the average squared difference between the observed values and the predicted values . It is a typical regression analysis prediction accuracy metric . The model's prediction accuracy is summarized in a single value , which makes the process of comparing and choosing models easier . By eliminating the subjectivity associated with human perception, MSE enables an objective method of evaluating video quality. It is hence beneficial for automated quality assessment systems.

The formula for MSE [48] is given in equation 4.1. For true labels in the data, regression model makes predictions . The difference between each true label and the corresponding predicted value is calculate and squared . All the squared errors are summed and divided by the total number of data points to get the MSE.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4.1}$$

Where,

$$\begin{aligned} \hat{y}_i &= \text{Predicted value for the ith observation} \\ y_i &= \text{Observed value or true label for the ith observation} \\ n &= \text{Sample size or total number of datapoints} \end{aligned}$$

Lower MSE indicates that the predicted values are closer to the actual values , suggesting a better fit of the model . Higher MSE indicates that the predicted values are farther from the actual values , suggesting a poorer fit of the model. Because the MSE is scale-dependent , the size of the data being forecasted affects its value . MSE values

are often higher for larger values in the dataset . In our case, since feature scaling is not applied , hence the values of MSE are justified.

4.3.2 PCC

The statistical measure known as Pearson's correlation coefficient is used to quantify the direction and strength of a linear relationship between two variables [49] . It has a range of -1 to 1 . There is no linear correlation when the value is 0 , perfect negative linear correlation is shown by a value of -1 , and perfect positive linear correlation is shown by a value of 1 . In our case, X is array of features and Y is their corresponding Mean Observer Score. Mean is calculate for both features and labels . For each data point , deviations are calculated from the mean values . Finally, PCC is computed. Formula for PCC [50] is given by equation 4.3.2.

$$\rho(X,Y) = \frac{\sum_{t=1}^{N} (X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^{N} (X_t - \bar{X})^2 \sum_{t=1}^{N} (Y_t - \bar{Y})^2}}$$

where

$$X_t \ , \ Y_t \ = \text{Individual sample points}$$
$$\bar{X}, \bar{Y} = \text{Means of (X) and (Y)}$$

A high Pearson correlation coefficient would suggest that the objective measurements are good at predicting how viewers will evaluate the quality of the video because they are closely correlated with human perceptions of video quality . On the other hand , a low Pearson correlation would imply that the objective metrics are not a good enough representation of the opinions of the viewers , which would mean that the methods of assessment need to be improved [51].

4.3.3 SROCC

A measure of the degree of relationship between two variables based on their ranks rather than their actual values is the Spearman rank correlation coefficient [52]. It helps to identify non linear relationship between features and labels . The Spearman rank correlation is less impacted by outliers since it employs ranks rather than actual values . It is particularly advantageous when dealing with non-normally distributed data, ordinal data, or data with outliers , making it a versatile choice for many statistical analyses . This indicates that when extreme values are available , it can offer a more trustworthy measure of correlation.

Features from the video frames and their corresponding mean observer scores are given as labels . labels and features are ranked in ascending order , such that if two or more values have same rank , then we can compute average rank for total number of values that have same rank. For each pair of ranks , difference is calculated. These differences

are squared and then summation is applied to these squared differences. Finally we can calculate SROCC using the equation 4.2

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{4.2}$$

Where,

$$\begin{aligned} \rho &= \text{Spearman rank correlation coefficient} \\ d_i &= \text{Difference between the ranks of each pair} \\ n &= \text{Number of pairs of ranks} \end{aligned}$$

The resulting value of SROCC will range from -1 to 1, where 1 indicates a perfect positive correlation, -1 indicates a perfect negative correlation, and 0 indicates no correlation. Values in between are assigned either weak, moderate or strong correlations. Usually correlation values less than 0.5 indicate weak, around 0.5 means a moderate, and values greater than or equal to 1 indicate strong correlation.

4.3.4 KRCC

A non-parametric method for determining the degree and direction of a link between two variables is the Kendall's tau correlation coefficient . It assesses the extent to which a monotonic function may adequately reflect the connection between two variables [53]. KRCC can capture both linear and non-linear relationships between features and labels . The features and labels are ranked in ascending order . Ordered pairs are formed from these ranks. After this, the Concordant and Discordant Pairs are counted . A pair of observations is concordant if the ranks for both observations follow the same order . On the other hand , a pair is discordant if the ranks for the observations follow different orders. The formula for the Kendall rank correlation coefficient [54] is given by:

$$\tau = 1 - \frac{2 \times [d_\Delta(P_1, P_2)]}{N(N-1)}] \tag{4.3}$$

where,

$$\begin{aligned} d_\Delta(P_1, P_2) &= \text{Symmetric difference distance between the two sets of ordered pairs} \\ N &= \text{Total Number of Concordant and Discordant pairs in the set} \end{aligned}$$

There are three possible values for the coefficient , -1 for a perfect negative correlation , +1 for a perfect positive correlation , and 0 for no correlation. In comparison to other correlation methods , it is especially helpful for ordinal data and less susceptible to outliers and is used to compare the subjective and objective evaluation findings of image/video quality.

4.4 Performance Analysis

4.4.1 PROPOSED METHODOLOGY 1- FEATURE EXTRACTION

4.4.1.1 Evaluation using FRIQUEE

FRIQUEE has shown better performance as compared to other feature extraction techniques since it consists of 560 features. Results of different models using FRIQUEE are discussed below. PCC measures the linear relationship between variables, whereas, SROCC and KRCC measures the strength and direction of association between two ranked variables.
Table 4.1 shows comparison of correlation coefficients [55] and MSE obtained by different regression models. Model can be evaluated using various metrics such as PCC, KRCC and SROCC, each providing insights into different aspects of the model's predictive capabilities.

Table 4.1: Comparison of correlation coefficients and MSE of models using features extracted by FRIQUEE

Model	MSE	PCC	KRCC	SROCC
Linear Regression	163.94	0.671	0.501	0.717
Decision Tree	170.78	0.685	0.404	0.571
Random Forest	114.48	**0.832**	0.631	**0.817**
ConvNet	109.09	**0.803**	0.521	0.669
Support Vector Regression	88.728	**0.856**	0.639	**0.831**

Since feature scaling is not applied, hence we can interpret the model performance using MSE by comparing MSE values of models with eachother. SVR has the lowest MSE of 88.728, therefore we can say that it is ranked first. ConvNet is ranked second with MSE of 109.09, random forest is ranked third with MSE of 114.48. Linear regression and Decision tree takes fourth and fifth rank, with MSE of 163.94 and 170.78 respectively. In Table 4.1, we can clearly see that Support Vector Regression performs better than rest of the models. PCC measures the linear relationship between the predicted values and the actual values. A PCC of 0.856 indicates a strong positive linear relationship. This suggests that the SVR model has good performance in terms of predicting values that are linearly related to the actual values. Kendall's tau measures the ordinal association between two measured quantities. It is less sensitive to the strength of the relationship compared to the Pearson correlation and focuses more on the ordinal (ranking) association. A KRCC of 0.639 indicates a fairly strong ordinal association. This implies that the model's predictions preserve the relative ordering of the data quite well, even if the exact values are not perfectly predicted. Spearman's rank correlation coefficient measures the strength and direction of the association between two ranked variables. Like Kendall's tau, it assesses how well the relationship between two variables can be described using a monotonic function. SROCC of 0.831 indicates a strong positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values effectively. We can interpret that SVR model with these correlation values is performing well, providing reliable and consistent predictions with respect to both linear and ordinal associations.

Random forest regression model is ranked second in performance after SVR. A PCC of 0.832 indicates a strong positive linear relationship. This suggests that the Random Forest Regression model has good performance in terms of predicting values that are linearly related to the actual values. KRCC value of 0.631 indicates a fairly strong ordinal association. This implies that the model's predictions preserve the relative ordering of the data quite well, even if the exact values are not perfectly predicted. SROCC of 0.817 indicates a strong positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values effectively. Random Forest Regression model has strong predictive performance, both in terms of linear accuracy and rank preservation. This is indicative of a robust model that not only predicts values closely aligned with the actual values but also maintains the correct order of those predictions.

ConvNet model has a good performance in capturing the linear relationship between predicted and actual values as indicated by its PCC of 0.803, though its performance is slightly lower compared to the SVR and Random Forest models. KRCC (0.521) and SROCC (0.669) indicate that the CNN model maintains a moderate to strong ordinal relationship between predictions and actual values. However, these values are also lower than those of the SVR and Random Forest models, suggesting that the CNN model is less effective in preserving the relative ordering of the data. This indicates that while the CNN can be a useful model, for this particular task, SVR and Random Forest might provide more reliable and consistent predictions.

Decision Tree Regression model has a reasonable performance in predicting values that are linearly related to the actual values, but its performance is weaker compared to the SVR, Random Forest, and CNN models as indicated by PCC of 0.685. It shows a moderate positive linear relationship. KRCC of 0.404 indicates a moderate ordinal association. This implies that the model's predictions preserve the relative ordering of the data to some extent, but not as effectively as the other models. SROCC of 0.571 indicates a moderate positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values to a moderate degree. Decision Tree Regression model with these correlation values is performing moderately well but is not as strong as the SVR, Random Forest, or CNN models in both linear accuracy and rank preservation.

Linear Regression model has a reasonable performance in predicting values that are linearly related to the actual values as seen by its PCC of 0.671, but its performance is weaker compared to the SVR, Random Forest, and CNN models. 0.501 indicates a moderate ordinal association. KRCC 0.501 implies that the model's predictions preserve the relative ordering of the data reasonably well. Whereas, SROCC of 0.717 indicates a strong positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values effectively.

4.4.1.2 Evaluation using OGIQA

Table 4.2 shows MSE and correlation coefficients obtained by KNN, Linear regression, Random forest regression and convolutional neural network. Since Normalization is not applied to the data, hence the values of MSE are greater in scale. However, we can interpret the results by comparing the values of MSE obtained by different models. ConvNet, with the lowest MSE of 145.37, is the best performing model among those compared. It effectively captures complex relationships in the data, leading to more accurate predictions. Random Forest Regression (MSE: 150.68) and Linear Regression (MSE: 152.41) have moderate performance, with Random Forest slightly outperforming Linear Regression. Random Forest's ensemble method gives it an edge over the simpler Linear Regression model. K-Nearest Neighbor (KNN) has the highest MSE (172.63), indicating the least effective performance for this regression task.

PCC captures linear trends in data whereas KRCC and SROCC assess the rank-order relationship between the predicted and actual values and are particularly useful when the relationship is monotonic but not necessarily linear. For correlation coefficients, values less than 0.3 indicate weak performance of the model, 0.3 to 0.7 indicate a moderate correlation and values greater than 0.7 indicate a very strong relationship between predicted and actual values.

Table 4.2: Comparison of correlation coefficients and MSE obtained by different models

Model	MSE	PCC	KRCC	SROCC
K Nearest Neighbor	172.63	0.679	0.544	0.711
Linear Regression	152.41	0.739	0.523	0.744
Random Forest Regression	150.68	0.721	0.518	0.701
ConvNet	145.37	0.793	0.588	0.797

K Nearest Neighbor has a reasonable performance in predicting values that are linearly related to the actual values as suggested by PCC of 0.679. KRCC (0.544) and SROCC (0.711) indicate that the model maintains the ordinal relationship between predictions and actual values to a good extent and is relatively good at preserving the rank order of the data. KNN model shows moderate performance overall.

Linear Regression model is a solid performer, particularly in capturing linear relationships and maintaining rank order. PCC of 0.739 indicates a strong positive linear relationship. This suggests that the Linear Regression model has good performance in predicting values that are linearly related to the actual values. KRCC of 0.523 indicates a moderate ordinal association. This implies that the model's predictions preserve the relative ordering of the data reasonably well. SROCC of 0.744 indicates a strong positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values effectively.

Random Forest Regression model captures the linear relationship between predicted and actual values quite well as indicated by PCC of 0.721. This strong positive correlation

indicates good predictive performance for linear relationships. KRCC of 0.518 indicates a moderate to strong ordinal association. This implies that the model's predictions preserve the relative ordering of the data fairly well. SROCC of 0.701 indicates a strong positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values effectively.

ConvNet model demonstrates exceptional performance across all correlation metrics, indicating its efficacy in capturing both linear and ordinal associations in the data. PCC of 0.793 indicates a strong positive linear relationship. This suggests that the CNN model has excellent performance in predicting values that are linearly related to the actual values. KRCC of 0.588 indicates a strong ordinal association. This implies that the model's predictions preserve the relative ordering of the data quite well. SROCC of 0.797 indicates a strong positive monotonic relationship, suggesting that the model's predictions maintain the rank order of the actual values very effectively. ConvNet excels in capturing complex patterns and relationships in the data, as evidenced by its high PCC, strong KRCC, and SROCC.

4.4.1.3 Evaluation using GWH-GLBP

Features obtained by GWH-GLBP are evalauted using support vector regression, decision tree, random forest, linear regression and ConvNet.
Table 4.3, shows comparison of evaluation metrics. The scale of MSE is influenced by the scale of the target variable. Feature scaling is not applied to the data. Therefore we can interpret the model performance, in terms of MSE by comparing with eachother. Linear Regression has the lowest MSE of 105.73, among the models, indicating the best performance in terms of minimizing prediction errors. This suggests that Linear Regression provides the most accurate predictions compared to the other models. Decision Tree Regression has the highest MSE (209.01) among the models compared, indicating the poorest performance in terms of minimizing prediction errors. This suggests that Decision Tree Regression is less effective for this particular regression task compared to the other models. Support Vector Regression has a higher MSE (134.89) compared to Linear Regression and slightly higher than Random Forest Regression (MSE: 132.85). This suggests that SVR performs moderately well but is not as accurate as Linear Regression. ConvNet has a similar MSE (131.34) to Random Forest Regression, indicating comparable performance in terms of minimizing prediction errors. However, it performs better than Decision Tree and Support Vector Regression.

Decision Tree Regression model shows moderate performance in capturing both linear and ordinal associations in the data. PCC of 0.571 indicates a moderate positive linear relationship. This suggests that the Decision Tree Regression model has a moderate performance in predicting values that are linearly related to the actual values. KRCC of 0.441 indicates a moderate ordinal association. This implies that the model's predictions preserve the relative ordering of the data to some extent. SROCC of 0.553 indicates a moderate positive monotonic relationship. This suggests that the model's predictions

Table 4.3: Comparison of correlation coefficients and MSE using features extracted by GWH-GLBP

Model	MSE	PCC	KRCC	SROCC
Decision Tree	209.01	0.571	0.441	0.553
Linear Regression	105.73	0.748	0.555	0.781
Random Forest Regression	132.85	0.758	0.528	0.713
Support Vector Regression	134.89	0.727	0.576	0.769
ConvNet	131.34	0.743	0.594	0.775

maintain the rank order of the actual values to a moderate degree.

Linear Regression model performs exceptionally well, with high correlation coefficients across all metrics, indicating superior predictive accuracy and rank preservation. PCC of 0.748 suggests that the model captures linear relationship between predicted and actual values quite well, indicating high predictive accuracy for linear relationships. KRCC of 0.555 indicates a moderate to strong ordinal association. This implies that the model's predictions preserve the relative ordering of the data quite well. SROCC of 0.781 indicates a strong positive monotonic relationship. This suggests that the model's predictions maintain the rank order of the actual values effectively indicating robust performance in capturing the overall trends in the data.

Random Forest Regression has a PCC of 0.758 which mean that the model exhibits good performance in demonstrating linear trends between actual and predicted values. On the other hand, KRCC has a value of 0.528 and SROCC has a value of 0.713, which indicate moderate to strong similarity in ranking orders, respectively.

4.4.1.4 Evaluation using CURVELET

Curvelet is evaluated on five regression models, KNN, Random forest, Decision tree, Linear regression, ConvNet and SVR. Table 4.4,shows correlation coefficients and MSE obtained by six regression models. Again, Linear regression has lowest MSE of 133.78, SVR is ranked second with MSE of 134.68, Random forest is ranked third with MSE of 139.32, ConvNet is ranked fourth with MSE of 141.48, KNN is ranked fifth with MSE 171.44 and finally Decision tree is ranked sixth with MSE of 179.4. For correlation parameters, SVR has outperformed with PCC 0.827, KRCC 0.613 and SROCC 0.808. ConvNet has PCC 0.782, KRCC 0.595 and SROCC 0.778. Linear regression has PCC 0.739, KRCC 0.563 and SROCC 0.741. Decision tree has a PCC of 0.666, KRCC 0.385 and SROCC of 0.523. Random forest has a PCC of 0.783, KRCC 0.597 and srocc OF 0.762. KNN has a PCC value of 0.656, KRCC 0.405 and SROCC of 0.541.

4.4.1.5 Evaluation using BRISQUE

Table 4.5 shows comparison of correlation coefficients and MSE obtained by different regression models that use features extracted using BRISQUE. As it can be seen

Table 4.4: Comparison of correlation coefficients and MSE using CURVELET features

Model	MSE	PCC	KRCC	SROCC
K Nearest Neighbor	171.44	0.656	0.405	0.541
Random Forest	139.32	0.783	0.597	0.762
Decision Tree	179.4	0.666	0.385	0.523
Linear Regression	133.78	0.739	0.563	0.741
ConvNet	141.48	0.782	0.595	0.778
Support Vector Regression	134.68	**0.827**	0.613	**0.808**

that SVR has lowest MSE of 116.45, Linear regression has second lowest MSE of 139.51, Random forest has third lowest MSE of 153.67, ConvNet has fourth lowest MSE of 171.17, Decision tree has fifth lowest MSE of 216.23. In terms of correlation parameters, Linear regression has shown top performance as compared to other models with a PCC of 0.767, KRCC 0.623 and SROCC of 0.813. Random forest has a PCC of 0.711, KRCC 0.542 and SROCC of 0.738. ConvNet has a PCC of 0.683, KRCC 0.539 and SROCC 0.763. SVR has a PCC of 0.681, KRCC 0.464 and SROCC 0.692. Decision tree has moderate performance with a PCC of 0.597, KRCC 0.397 and SROCC 0.574.

Table 4.5: Comparison of correlation coefficients and MSE using BRISQUE features

Model	MSE	PCC	KRCC	SROCC
Random Forest	153.67	0.711	0.542	0.738
Support Vector Regression	116.45	0.681	0.464	0.692
Decision Tree	216.23	0.597	0.397	0.574
ConvNet	171.17	0.683	0.539	0.763
Linear Regression	139.51	0.767	0.623	**0.813**

We present a consolidated table 4.6 which summarizes the results of all feature extraction techniques for our proposed methodology 1- Feature Extraction.

Table 4.6: Summary of results obtained by different feature extraction techniques for Proposed Methodology 1- Feature Extraction

Feature Extraction Techniques	MSE	PCC	KRCC	SROCC
OGIQA	145.37	0.793	0.588	0.797
BRISQUE	139.51	0.767	0.623	0.813
CURVELET	134.68	0.827	0.613	0.808
GWH-GLBP	105.73	0.748	0.555	0.781
FRIQUEE	**88.728**	**0.856**	**0.639**	**0.831**

4.4.2 PROPOSED METHODOLOGY 2- FEATURE SELECTION

4.4.2.1 Feature Selection for FRIQUEE

Table 4.7 shows a summary of correlation coefficients and MSE obtained by different regression models. Genetic search algorithm is used to obtain selective features.

Table 4.7: Comparison of correlation coefficients and MSE using attribute selection for FRIQUEE

Model	MSE	PCC	KRCC	SROCC
Support Vector Regression	115.45	0.752	0.510	0.701
Random Forest	**2.530**	**0.995**	**0.961**	**0.993**
Linear Regression	18.758	**0.963**	**0.821**	**0.953**
Decision Tree	9.814	**0.981**	**0.968**	**0.981**
ConvNet	**1.318**	**0.997**	**0.955**	**0.995**

According to Table 4.7,ConvNet model's performance metrics collectively suggest that the model is performing exceptionally well. The low MSE of 1.318 indicates accurate predictions, while the high PCC 0.997, KRCC 0.955, and SROCC 0.995 show that the model captures the true relationships and ranking within the data very well. These high correlation values across different metrics (PCC, KRCC, SROCC) suggest that the model is robust and reliable in different aspects of prediction, both in terms of linear relationship and rank-order consistency. While low MSE indicates that the model is making predictions that are very close to the actual values, suggesting high accuracy.

Decision tree has higher MSE of 9.814 as compared to the ConvNet model. This indicates that the decision tree has larger average prediction errors. Nonetheless, the MSE of 9.814 still suggests reasonably accurate predictions. The high PCC 0.981, KRCC 0.968, and SROCC 0.981, suggest that the model captures true relationships within the data well, preserving both linear relationships and rank orders effectively.

The performance metrics of the random forest model indicate very strong performance, with particularly notable accuracy and consistency. A MSE of 2.530 suggests that the model's predictions are quite close to the actual values. PCC value of 0.995 suggests an almost perfect positive linear relationship. This high correlation indicates that the random forest's predictions are very closely aligned with the true values. KRCC value of 0.961 suggests a very strong agreement in the ranking of the predicted and actual values. SROCC value of 0.993 signifies an almost perfect monotonic relationship, indicating that the model's predictions maintain a very high degree of consistency with the ranks of the true values.

The performance metrics of the linear regression model indicate moderate performance, with less accuracy and consistency compared to more complex models like CNNs and random forest. A MSE of 18.758 suggests that the model's predictions have relatively large

average errors compared to the true values. PCC value of 0.963 indicates a strong positive linear relationship. This suggests that while there is a good linear fit, it is not as strong as that achieved by the more complex models. Kendall tau and Spearman rank correlation coefficients (0.821 and 0.953, respectively) indicate good but not perfect consistency in preserving the order and ranking of the data points. These values are lower than those of the other models, suggesting that the linear regression model is less reliable in maintaining rank-order relationships.

Support Vector regression has a high MSE of 115.45 which reflects significant prediction errors, indicating low accuracy. PCC of 0.752 indicates a moderate linear relationship between predicted and actual values, showing that the model does not fit the data well linearly. The Kendall tau and Spearman rank correlation coefficients (0.510 and 0.701, respectively) indicate moderate consistency in preserving the order and ranking of the data points. These values are significantly lower than those of the other models, suggesting that the SVR model is less reliable in maintaining rank-order relationships.

4.4.2.2 Feature Selection for OGIQA

Important feature columns are selected for OGIQA using genetic search algorithm and results are re-evaluated as shown in Table 4.8.

The feature scaling is not applied hence values of MSE are larger in scale. But we can interpret the model performance by comparing the MSE of models with each other. The KNN (MSE 79.337), model shows best performance in terms of prediction accuracy, closely followed by the Random Forest model (MSE 80.183). Both of these models are better suited to the data at hand, capturing non-linear relationships and interactions effectively. On the other hand, Linear Regression (MSE 183.66) and CNN (MSE 182.22) perform poorly, likely due to their inability to handle the complexity and specific nature of the data without feature selection.

Table 4.8: Comparison of correlation coefficients and MSE using attribute selection for OGIQA

Model	MSE	PCC	KRCC	SROCC
Linear Regression	183.66	0.542	0.354	0.526
K Nearest Neighbor	79.337	**0.836**	0.695	**0.831**
Random Forest	80.183	**0.832**	0.678	**0.825**
ConvNet	182.22	0.543	0.349	0.515

The KNN model demonstrates strong ability in capturing the overall trend and ranking of the data, as indicated by the high PCC and SROCC with values of 0.836 and 0.831 respectively. KNN model is capturing the overall trend and rank ordering in the data very well. This means that the model is good at predicting the relative ordering of the data points. The KRCC(0.695) also shows a moderately strong ordinal association, reinforcing that the model is reasonably good at rank ordering.

In case of Random forest, the high PCC (0.832) and SROCC (0.825) indicate that the model is good at capturing the overall trend and rank ordering in the data, showing strong linear and monotonic relationships between the predicted and actual values. The Kendall rank correlation coefficient (0.678) indicates a moderately strong ordinal association, reinforcing that the model is fairly good at maintaining the correct order of data points, although not perfectly.

Given that ConvNet is typically designed for spatial or image data, their application to non-image/tabular data might not be optimal without significant adjustments. PCC value of 0.543 indicates a moderate positive linear relationship. This suggests that while there is some degree of linear alignment between the ConvNet's predictions and the actual values, the relationship is not particularly strong. KRCC value of 0.349 indicates a weak to moderate positive correlation in terms of rank ordering. This means that the CNN's predictions somewhat preserve the order of the data points but not very reliably. This implies that the model's ability to predict the correct order or ranking of data points is limited. SROCC value of 0.515 indicates a moderate positive monotonic relationship. This suggests that there is some degree of consistency in the rank order of the predictions, but it is not particularly strong.

Given that Linear Regression assumes a linear relationship between features and target, it might not be well-suited for capturing complex patterns in the data. The PCC of 0.542 indicates that Linear Regression has a moderate ability to capture linear trends in the data. However, this relationship is not strong enough to suggest high accuracy. Both KRCC 0.354 and SROCC 0.526 are relatively low, indicating that the model struggles to preserve the correct rank order of the data points.

4.4.2.3 Feature Selection for GWH-GLBP

GWH-GLBP features are re-evaluated after feature selection using genetic algorithm. It can be seen that results have improved. Table 4.9 shows summary of the results.

Table 4.9: Comparison of correlation coefficients and MSE using attribute selection for GWH-GLBP

Model	MSE	PCC	KRCC	SROCC
Support Vector Regression	108.31	0.767	0.544	0.745
Random Forest	7.065	**0.986**	**0.938**	**0.985**
Linear Regression	117.7	0.741	0.523	0.731
Decision Tree	17.03	**0.967**	**0.942**	**0.969**

Random Forest Regression has the lowest MSE among the models, indicating it performs the best in terms of prediction accuracy. This model is highly effective at capturing the underlying patterns in the data, significantly reducing the prediction errors compared to other models. Random Forests are robust to over fitting due to their ensemble nature and

are capable of handling complex, non-linear relationships in the data. Both SVR (108.31) and Linear Regression (117.7) show high MSE values, indicating poor model performance. Decision Tree, while better than SVR and Linear Regression, still has a significantly higher MSE (17.03) compared to Random Forest. This suggests it is capturing some patterns but not as effectively as the ensemble method.

The Random Forest model is extremely effective at capturing the linear trends in the data, leading to highly accurate predictions in terms of both direction and magnitude. PCC value of 0.986 indicates an exceptionally strong positive linear relationship. This suggests that the predicted values from the Random Forest model are almost perfectly aligned with the actual values. KRCC value of 0.938 and SROCC 0.985 indicates a very strong positive correlation in terms of rank ordering. The model is excellent at maintaining the correct rank order of the predictions. Overall it implies that Random Forest Regression model is well-tuned and highly effective for the given dataset.

Decision tree regression results indicate robust performance. PCC (0.967) indicates that the model captures the linear relationships in the data very well, leading to highly accurate predictions. Both KRCC (0.942) and SROCC (0.969) are very high, indicating that the model is excellent at preserving the rank order of the data points. The high correlation metrics across all three coefficients demonstrate that the Decision Tree model performs exceptionally well in capturing both linear relationships and the correct ranking of data points.

Support Vector regression shows a strong degree of accuracy in its predictions, but not as high as the best-performing models (like Random Forest or Decision Tree). The PCC (0.767) indicates that the SVR model captures the linear relationships in the data fairly well, resulting in relatively accurate predictions. KRCC value of 0.544 indicates a moderate positive correlation in terms of rank ordering. This means that the SVR model somewhat preserves the order of the data points, but not very reliably. SROCC value of 0.745 indicates a strong positive monotonic relationship. This suggests that the rank order of the predicted values reasonably matches the rank order of the actual values.

The PCC (0.741) indicates that the Linear Regression model captures a significant portion of the linear relationships in the data, resulting in moderately accurate predictions. Both KRCC (0.523) and SROCC (0.731) are moderate, indicating that while the model preserves the rank order of the data points to some extent, there is room for improvement.

4.4.2.4 Feature Selection for CURVELET

Table 4.10, shows summary of results obtained after re-evaluating the models with selected features.

Comparing the mean squared error (MSE) across different models can provide insights into their relative performance in terms of prediction accuracy. Based on MSE values, it can be interpreted that Random Forest regression with MSE of 76.03 performs the best

Table 4.10: Comparison of correlation coefficients and MSE using attribute selection for CURVELET

Model	MSE	PCC	KRCC	SROCC
K Nearest Neighbor	109.10	0.767	0.597	0.748
Random Forest	76.03	**0.842**	0.689	**0.828**
Decision Tree	123.9	0.763	0.671	0.749
Linear Regression	165.88	0.604	0.404	0.575
ConvNet	132.98	0.707	0.495	0.675
Support Vector Regression	170.54	0.611	0.434	0.588

among the models, followed by K Nearest Neighbors (MSE 109.10), Convolutional Neural Network (MSE 132.98), Decision Tree (MSE 123.9), Linear Regression (MSE 165.88), and Support Vector Regression (MSE 170.54), in descending order of performance.

K Nearest Neighbor shows high values of all three correlation coefficients (PCC 0.767, KRCC 0.597, and SROCC 0.748) indicate that the predicted values are closely related to the actual values, with strong positive relationships in terms of both linear and rank-based associations. This suggests that the K Nearest Neighbor model is performing well in capturing the underlying patterns in the data and providing accurate predictions.

Random Forest regression has a PCC of 0.842. This indicates a strong positive linear relationship between the predicted and actual values generated by the model. KRCC of 0.689 suggests a moderately strong positive association between the predicted and actual values based on their ranks. While slightly lower than the Pearson correlation coefficient, it still indicates a strong tendency for the predicted and actual values to move together. SROCC value of 0.828 suggests a strong positive association between the predicted and actual values. Overall, Random Forest Regression model is performing well in capturing the underlying patterns in the data and providing accurate predictions.

PCC of 0.763 indicates a strong positive linear relationship between the predicted and actual values generated by Decision Tree Regression. KRCC of 0.671 indicates a moderately strong positive association between the predicted and actual values based on their ranks. While slightly lower than the Pearson correlation coefficient, it still implies a robust relationship between the predicted and actual values. Similar to KRCC, SROCC value of 0.749, measures the strength and direction of the association between two variables based on their ranks. SROCC shows a strong positive association between the predicted and actual values, which is slightly lower than the PCC but still indicates a strong relationship.

Overall, the correlation coefficients for Linear Regression are lower compared to the previous models, indicating a somewhat weaker relationship between the predicted and actual values. This suggests that the linear regression model might not capture the underlying patterns in the data as effectively as more complex models like Random Forest or KNN. However, it still provides valuable predictive insights, especially if the relationship between the variables is approximately linear. PCC of 0.604 indicates a moderate positive

linear relationship between the predicted and actual values generated by the Linear Regression. KRCC of 0.404 indicates a moderate positive association between the predicted and actual values based on their ranks. SROCC value of 0.575 suggests a moderate positive association between the predicted and actual values, indicating that the linear regression model captures some but not all of the underlying patterns in the data.

ConvNet effectively captures some of the underlying patterns in the data, especially considering its ability to handle complex spatial relationships in data such as images. However, there might still be room for improvement, as evidenced by the correlation coefficients not being extremely high. PCC of 0.707 indicates a moderately strong positive linear relationship between the predicted and actual values. KRCC of 0.495 and SROCC 0.675 suggests a moderate positive association between the predicted and actual values based on their ranks. Overall, the correlation coefficients for the ConvNet indicate a moderately strong relationship between the predicted and actual values, both in terms of linear and rank-based associations.

Overall, the correlation coefficients for the Support Vector Regression indicate a moderately strong relationship between the predicted and actual values, both in terms of linear and rank-based associations. This suggests that the Support Vector Regression effectively captures some of the underlying patterns in the data. However, like with other models, there might still be room for improvement, as evidenced by the correlation coefficients not being extremely high. It has a PCC, KRCC and SROCC values of 0.611, 0.434 and 0.588 respectively.

4.4.2.5 Feature Selection for BRISQUE

Feature selection for BRISQUE features was performed using genetic search algorithm. Table 4.11 shows performance metrics obtained by different regression models.

Table 4.11: Comparison of correlation coefficients and MSE using attribute selection for BRISQUE

Model	MSE	PCC	KRCC	SROCC
Random Forest	24.32	**0.953**	**0.866**	**0.951**
Support Vector Regression	144.52	0.679	0.516	0.685
Decision Tree	63.385	**0.878**	**0.850**	**0.887**
ConvNet	47.855	**0.916**	0.761	**0.914**
Linear Regression	144.61	0.668	0.470	0.674

Mean squared errors of models are compared with eachother since feature scaling is not applied. Random forest has the lowest MSE of 24.32, among the models compared. This indicates that it captures the underlying patterns in the data very well. The ensemble nature of random forest, which combines multiple decision trees, allows it to model complex relationships and interactions between features effectively. The ConvNet has a relatively low MSE of 47.855, though not as low as the random forest. This suggests that

while the ConvNet is good at capturing complex patterns, it might not be as optimal as the random forest for this particular dataset. ConvNets are particularly strong in handling spatial data like images. Decision Tree (MSE 63.385) performs moderately well but not as well as its ensemble counterpart (random forest), highlighting the advantage of ensemble methods. Support Vector Regression has MSE (144.52) comparable to linear regression (MSE 144.61), suggesting that it struggles to capture the underlying patterns in the data. SVR can be effective with the right kernel and parameters, but it might not be well-suited for this dataset in its current configuration.

Correlations close to 1 suggests that the model's predictions are highly linearly correlated with the actual values, meaning that as one set of values increases, the other set also increases. PCC of 0.953 indicates a very strong positive linear relationship between the predicted and actual values generated by the Random Forest Regression model. KRCC value of 0.866 suggests that the relative rankings of the predicted values closely match the rankings of the actual values. SROCC of 0.951 also indicates a very strong positive relationship between the predicted and actual values based on their ranks. Random Forest Regression model has an excellent performance as indicated by its high correlation coefficients. It captures both the linear and ordinal relationships between the predicted and actual values very effectively. The high correlation coefficients imply that the model predictions are very reliable and closely match the actual outcomes.

Support Vector Regression shows moderate performance. It has PCC of 0.679. This value indicates a moderately strong positive linear relationship between the predicted and actual values. While there is a significant linear association, a considerable amount of variance in the actual values is not captured by the model's predictions. KRCC value of 0.516 suggests that there is a moderate level of agreement in the relative ordering of the predicted and actual values, indicating that the model correctly ranks the predictions to some extent but with considerable mismatches. SROCC value of 0.685 suggests that there is a fairly strong monotonic relationship, indicating that as one set of values increases, so does the other, but not perfectly.

Decision Tree Regression model is performing excellently based on the given correlation coefficients, indicating it is a strong choice for the dataset in question. The PCC of 0.878 suggests that the Decision Tree Regression model has a very strong ability to predict the actual values in a linear context. The high KRCC (0.850) and SROCC (0.887) coefficients indicate that the model accurately preserves the rank order of the data. This implies that the model not only captures linear relationships but also correctly ranks the predictions relative to the actual values.

ConvNet has shown excellent performance both in terms of high predictive accuracy and robust rank preservation. A PCC of 0.916 suggests that the predicted values closely follow the actual values in a linear manner. This high value implies that the CNN model is highly effective in capturing the linear patterns within the data, making it very reliable for prediction. KRCC value of 0.761 suggests a strong level of agreement in the ordering of

the predicted and actual values, indicating that the ConvNet model ranks the predictions similarly to the actual values with a high degree of consistency. SROCC value of 0.914 implies that the predictions and actual values have a near-perfect monotonic relationship, meaning that as one set of values increases, the other set increases as well, with very few exceptions.

Linear Regression model shows moderate performance. PCC of 0.668 indicates a moderately strong linear relationship between the predicted and actual values, implying that the model captures some of the linear patterns in the data but leaves a substantial amount unexplained. KRCC 0.470): Suggests a moderate level of agreement in the rank order of the predicted and actual values. SROCC of 0.674 implies a moderately strong monotonic relationship, indicating that the model predictions increase with the actual values, but with noticeable discrepancies.

We summarize our results from the above experimentation in a consolidated table 4.12, for our proposed methodology 2- Feature Selection.

Table 4.12: Summary of results obtained by different feature extraction techniques for Proposed Methodology 2- Feature Selection

Feature Extraction Techniques	**MSE**	**PCC**	**KRCC**	**SROCC**
OGIQA	79.337	0.836	0.695	0.831
BRISQUE	24.32	0.953	0.866	0.951
CURVELET	76.03	0.842	0.689	0.828
GWH-GLBP	7.065	0.986	0.938	0.985
FRIQUEE	**1.318**	**0.997**	**0.955**	**0.995**

4.4.3 PROPOSED METHODOLOGY 3- VISUAL SALIENCY

4.4.3.1 Performance Of Visual Saliency Features By FRIQUEE

In Table 4.13, correlation coefficients and MSE obtained by different regression models are compared. Since feature scaling is not applied, therefore MSE of models are compared with eachother.
Random Forest Regression has the lowest MSE of 151.9, which means it is performing well as compared to other models. It serves raw data well, which could be attributed to its nature as an averaging technique, as an ensemble model. SVR has a MSE of 157.50 and it is capable to determine a good hyperplane which minimizes the error for the current data set. ConvNets are capable of processing complex patterns It has a MSE of 173.3 which translates into the fact that this model is better than linear regression and decision tree regression models. The Linear Regression method has the Mean Squared Error of 184.9, The accuracy of it is higher than that of the decision tree, but it is not the highest. Decision

Tree has a maximum Mean Squared Error of 262.8 in this case that makes the decision tree regression model the worst, in this situation among all other models.

Table 4.13: Comparison of Evaluation Parameters Using Visual Saliency Features Extracted by FRIQUEE

Model	MSE	PCC	KRCC	SROCC
Random Forest	151.90	0.724	0.515	0.686
Support Vector Regression	157.50	0.748	0.546	0.758
Decision Tree	262.8	0.572	0.446	0.609
ConvNet	173.3	0.682	0.508	0.709
Linear Regression	184.9	0.755	0.616	0.792

Random Forest has a PCC of 0.724 which indicates that the prediction models have reasonably high levels of accuracy in identifying linear movement. The KRCC describes the ordinal relation between two measured values. It measures the ability of the monotonic function to describe the relationship between the predicted and actual values. A value of 0.515 implies a moderate to strong positive monotonic relationship. This mean that one variable increases or decrease in the same proportion as the other variable, but not in a direct proportion. Spearman rank correlation coefficients indicate the magnitude and sign of the association between two ranked variables. SROCC is a parametric measure that explains the extent to which the actual function between the predicted values and the actual values can be described by a monotonically increasing or decreasing function. A value of 0.686 is an evidence of positive straight line relationship. This means that there is a high degree of similarity between the order of the predicted values and the actual values.

According to the given correlation metrics, it is observed that support vector regression model performs better than random forest model in terms of its performance. PCC are also utilized to gauge extent of linearity between predicted and observed values. A value of 0.748 means a strong straight-line relationship which is stronger than 0.724 for the random forest model. Therefore, SVR captures linear trends in data well where as predicted values increase with actual ones. On the other hand, KRCC measures the strength of monotonic relation between two variables measured on an ordinal scale. KRCC value of 0.546 indicates a moderate to strong monotonic relationship, slightly stronger than the random forest model with a value of 0.515. As such, this shows that SVR retains a good consistency in ranking order among predicted and actual as well as SROCC measures monotonic association between two ranked variables. SROCC value of 0.758 suggests a strong positive monotonic relationship, stronger than the random forest model.

For Decision tree, PCC value of 0.572 signifies a moderate and positive linear relationship, which implies that there is some linear relationship but it's not as strong as the random forest (0.724) and SVR (0.748) models. The moderate linear relationship shows that for the decision tree, it does capture some linear trends in the data but not as well as other models do. KRCC value of 0.446 indicates a moderate monotonic relationship, this is weaker

than the random forest (0.515) and SVR (0.546) models. The moderate KRCC implies that in decision tree, there exists some consistency in how predicted values are ranked compared to actual values, but not quite like other models would be. A SROCC value of 0.609 means a moderate to strong positive monotonicity, weaker than for Random Forests (0.686), SVRs(0.758). The least correlation between classes suggests that a decision tree generally preserves the rank ordering of predicted and actual values reasonably well though not quite so much when compared with other models.

ConvNet model exhibits good performance by displaying strong linear correlations as well as strong monotonic relationships. PCC value of 0.682 indicates a very high positive linear relationship. It therefore means that the ConvNet model captures linear trends in the data quite well, although less effectively than random forest and SVR models do. KRCC value of 0.508 shows a moderate to strong monotonic relationship which is better than decision tree (0.446) and slightly weaker than both random forest (0.515) and SVR (0.546). This implies that the ConvNet model maintains a fairly good consistency in rank ordering of predicted and actual values by SROCC 0.709, which is better than Decision Tree's 0.609 but slightly weaker than Random Forest's 0.686 and SVR's 0.758. Therefore, it can be deduced that the ConvNet model does a pretty good job at keeping rank order preservation between predicted and actual outputs; hence its performance could be said to be satisfactory due to this reason alone.

Linear regression is a simple and interpretable model. Its strong performance, as indicated by the correlation metrics, makes it an attractive choice, especially if the data has a mainly linear relationship. The correlation between predictions and actual values in the linear regression model is very strong (PCC 0.755), which shows that this model is highly capable of identifying linear trends. Thus, it has good predictive power for capturing linear trends. It surpasses all other models on this aspect too. Also, KRCC (0.616) and SROCC (0.792) show that the model also exhibits very strong monotonic relationships than any other models examined in this study. This implies that compared to other models, linear regression is effective in preserving overall rankings of predicted outcomes with respect to observed ones.

4.4.3.2 Performance Of Visual Saliency Features By BRISQUE

Again feature scaling is not applied hence we can compare the MSE of models with eachother. In Table 4.14, among the models compared, ConvNet is the most accurate because it minimizes mean square error (MSE 129.3). The least accurate model indicated by the highest MSE is Decision Tree Regression (MSE 280.7). Performance differences indicate that complex models like ConvNet, capture patterns in data more effectively than Decision Trees and even some ensemble methods such as Random Forests (MSE 180.30), which are simpler. Support Vector Regression performs closely to CNN, implying it is a robust model for minimizing prediction error though not the best model. It has a MSE of 132.40. Linear Regression is third in performance, with a MSE of (150.3).

Table 4.14: Comparison of Evaluation Parameters Using Visual Saliency Features Extracted by BRISQUE

Model	MSE	PCC	KRCC	SROCC
Random Forest	180.30	0.633	0.479	0.652
Support Vector Regression	132.40	0.803	0.601	0.798
Decision Tree	280.7	0.461	0.35	0.446
ConvNet	129.3	0.775	0.614	0.795
Linear Regression	150.3	0.718	0.512	0.706

The Random Forest Regression model has a moderate to strong relationship between the predicted and actual values as shown by the coefficients. PCC value of 0.633 shows that the model captures a significant fraction of the linear relationship, while KRCC 0.479 and SROCC 0.652 point out that there is also some preservation of rank order in the model. The model performs reasonably well. However, there are still areas where improvements can be made. These correlations are not extremely high (closer to 1), which means that refining it could enable capturing relationships better in data sets.

Support Vector Regression models have high patterns of correlation coefficients indicating accurate predictions, and the order of the data preservation. Support vector regression (SVR) model has a great relationship between the predicted and actual values. PCC value of 0.803 suggests that the model covers the linear correlation of the relationship to a remarkable extent, while the KRCC (0.601) and SROCC (0.798) indicate that the model as well maintains the rank order of the values excellently. The high correlation values, particularly PCC and SROCC, give the feeling that the model is well performing regarding predictability and preservation of observance order.

Decision tree regression shows a correlation which is reasonable. The PCC of 0.461 indicates a moderate positive linear relation between the predicted values and actuals. A moderate positive monotonous relationships between predicted and actual values is indicated by (KRCC = 0.35) and (SROCC 0.446), with a slightly higher value for the latter one.

As for ConvNet, it is highly correlated both in linear and monotonicity: CNN model against the actual values. This tends to indicate analogous predictions of the model with the real numbers in various dimensions of correlation. The strong positive linear relation between predicted values and actual values is shown by (PCC 0.775), (KRCC 0.614) and (SROCC 0.795).

Linear Regression shows almost same result with a slight variation. PCC 0.718, KRCC 0.512 and SROCC 0.706, indicate same performance as ConvNet.

4.4.3.3 Performance Of Visual Saliency Features By GWH-GLBP

Table 4.15 shows performance comparison of different regression models on the basis of MSE, PCC, KRCC and SROCC. Uptil now, general guess about the results can be

made. GWH-GLBP shows similar results as Table 4.14. MSE of models can be compared with eachother. Linear regression has lowest MSE of (113.8). Random forest stands second in performance with a MSE of (128.1). On third, Support Vector Regression has shown a MSE of (147.1). ConvNet (MSE 192.6) and Decision Tree (MSE 299.5) are ranked fourth and fifth, respectively.
In terms of correlation coefficients, Linear regression has the highest correlation values. PCC value of 0.808, KRCC value of 0.616 and SROCC value of 0.797, all indicate that the model is performing best, as compared to other models in table. ConvNet has second highest correlations, with PCC value of 0.773, KRCC 0.535 and SROCC value of 0.736. Support vector has almost same results as ConvNet, with a PCC of 0.723, KRCC value of 0.545 and SROCC value of 0.752. Random Forest shows moderate performance with PCC value of 0.618, KRCC value of 0.427 and SROCC value of 0.596. Weakest performing model as seen in the Table 4.15 is Decision Tree, with PCC value of 0.369, KRCC value of 0.299 and SROCC value of 0.397.

Table 4.15: Comparison of Evaluation Parameters Using Visual Saliency Features Extracted by GWH-GLBP

Model	MSE	PCC	KRCC	SROCC
Decision Tree	299.5	0.369	0.299	0.397
Linear Regression	113.8	0.808	0.616	0.797
Random Forest	128.1	0.618	0.427	0.596
Support Vector Regression	147.1	0.723	0.545	0.752
ConvNet	192.6	0.773	0.535	0.736

4.4.3.4 Performance Of Visual Saliency Features By OGIQA

A glimpse of results for OGIQA indicate that performance of all regression models is poor. This is because OGIQA extracts only six features per frame. Linear Regression and ConvNet have almost same MSE with slight variation. Whereas, K Nearest Neighbor and Random Forest have same MSE with negligible difference.
K Nearest Neighbor has a slightly higher correlations as compared to other models. PCC value of 0.419, KRCC 0.318 and SROCC 0.463. Linear Regression has a PCC value of 0.403, KRCC value of 0.266 and SROCC value of 0.371. Random Forest has a PCC value of 0.377, KRCC value of 0.28, and SROCC value of 0.41. Finally, ConvNet has a PCC value of 0.418, KRCC value of 0.294 and SROCC value of 0.429.

After the above experimentation, We summarize our results in a consolidated table 4.17, for our proposed methodology 3- Visual Saliency.

Table 4.16: Comparison of Evaluation Parameters Using Visual Saliency Features Extracted by OGIQA

Model	MSE	PCC	KRCC	SROCC
K Nearest Neighbor	265.2	0.419	0.318	0.463
Linear Regression	244.1	0.403	0.266	0.371
Random Forest	258.3	0.377	0.28	0.41
ConvNet	246.4	0.418	0.294	0.429

Table 4.17: Summary of results obtained by different feature extraction techniques for Proposed Methodology 3- Visual Saliency

Feature Extraction Techniques	MSE	PCC	KRCC	SROCC
OGIQA	265.2	0.419	0.318	0.463
BRISQUE	132.40	0.803	0.601	0.798
FRIQUEE	184.9	0.755	0.616	0.792
GWH-GLBP	**113.8**	**0.808**	**0.616**	**0.797**

4.4.4 PROPOSED METHODOLOGY 4- VISUAL SALIENCY, FEATURE EXTRACTION COMBINED WITH FEATURE SELECTION

4.4.4.1 Re-Evaluating FRIQUEE With Attribute Selection

After performing attribute selection for visual saliency features extracted by FRIQUEE, it is evident from Table 4.18 that evaluation parameters have drastically improved. Again feature scaling is not applied but MSE of Random Forest model has decreased from 151 to 17.80, which is an exceptional improvement. PCC value of 0.967, KRCC 0.868 and SROCC 0.962, indicate that the model performance has improved to almost perfection. Decision Tree is second best performing model. Its MSE has also decreased from 262.8 to 56.612. As for correlation coefficients, moderate values have improved to stronger values. PCC value is 0.891, PCC value is 0.838 and SROCC value is 0.891. MSE of Linear regression has decreased from 184.9 to 60.61. Correlation parameters have also improved with a PCC of 0.876, KRCC 0.686 and SROCC 0.872. However, in case of Support Vector Regression, results have not improved rather they have increased the MSE to 179.99, correlation parameters have also decreased. It has PCC value of 0.559, KRCC value of 0.377 and SROCC value 0.541.

Table 4.18: Comparison of Evaluation Parameters Using Attribute Selection and Visual Saliency Features Extracted by FRIQUEE

Model	MSE	PCC	KRCC	SROCC
Random Forest	17.80	**0.967**	**0.868**	**0.962**
Linear Regression	60.61	**0.876**	0.686	**0.872**
Decision Tree	56.612	**0.891**	**0.838**	**0.891**
Support Vector Regression	179.99	0.559	0.377	0.541

4.4.4.2 Re-Evaluating BRISQUE With Attribute Selection

Overall results for BRISQUE have also improved with attribute selection, as it can be seen in Table 4.19.
Random Forest MSE has reduced from 180.30 to 68.99, which is a significant improvement. As for PCC, KRCC and SROCC, the values have also increased to 0.858, 0.717 and 0.857 respectively.
Decision tree has also shown improvement in results with a significantly lower MSE of 142.98. PCC value has increased from 0.461 to 0.727. KRCC has also increased from 0.35 to 0.671. Similarly, SROCC has increased from 0.446 to 0.74.
In case of Linear Regression and Support Vector Regression, performance of these models have decreased with attribute selection, both with higher MSE and correlation parameters.

Table 4.19: Comparison of Evaluation Parameters Using Attribute Selection and Visual Saliency Features Extracted by BRISQUE

Model	MSE	PCC	KRCC	SROCC
Random Forest	68.99	**0.858**	0.717	**0.857**
Support Vector Regression	167.30	0.609	0.456	0.614
Decision Tree	142.98	0.727	0.671	0.74
Linear Regression	184.76	0.54	0.386	0.575

4.4.4.3 Re-Evaluating GWH-GLBP With Attribute Selection

It can be seen in Table 4.20 that results for Decision Tree and Random Forest have improved. Both show decreased MSE value of 109.67 and 57.37, which is a massive improvement. Similar is the case of correlation parameters. Random Forest has improved PCC value of 0.883, KRCC value of 0.744 and SROCC value of 0.882. Decision Tree has PCC value of 0.789, KRCC value of 0.693 and SROCC 0.795.
However, Linear Regression and Support Vector Regression have increased MSE which shows poor performance. Also, a significant drop in correlation parameters for both models can be observed, as compared to previous results.

Table 4.20: Comparison of Evaluation Parameters Using Attribute Selection and Visual Saliency Features Extracted by GWH-GLBP

Model	MSE	PCC	KRCC	SROCC
Decision Tree	109.67	0.789	0.693	0.795
Linear Regression	162.52	0.614	0.438	0.642
Random Forest	57.37	**0.883**	0.744	**0.882**
Support Vector Regression	160.44	0.624	0.443	0.639

4.4.4.4 Re-Evaluating OGIQA With Attribute Selection

OGIQA results were poor overall in previous approach. After performing attribute selection for OGIQA visual saliency features, the results have decreased for all models. Table 4.21 shows the MSE and correlation parameters (PCC, KRCC and SROCC) for three different regression models.

Table 4.21: Comparison of Evaluation Parameters Using Attribute Selection and Visual Saliency Features Extracted by OGIQA

Model	MSE	PCC	KRCC	SROCC
K Nearest Neighbor	281.07	0.199	0.112	0.163
Linear Regression	246.19	0.239	0.134	0.196
Random Forest	272.54	0.219	0.124	0.182

After above experimentation, We summarize our results in a consolidated table 4.22 for our proposed methodology 4.

Table 4.22: Summary of results obtained by different feature extraction techniques for Proposed Methodology 4- Visual Saliency, Feature Extraction combined with Feature Selection

Feature Extraction Techniques	MSE	PCC	KRCC	SROCC
OGIQA	246.19	0.239	0.134	0.196
BRISQUE	68.99	0.858	0.717	0.857
FRIQUEE	**17.80**	**0.967**	**0.868**	**0.962**
GWH-GLBP	57.37	0.883	0.744	0.882

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Overview

This Chapter focuses on overall performance analysis and future recommendations.

5.2 Performance Evaluation of Proposed Methodology

Table 5.1 shows comparison of our proposed methodology with state of the art techniques. We can see that our proposed methodology shows better performance as compared to other techniques with higher PCC , KRCC and SROCC. This indicates that identifying the video's most significant areas for perceived quality is made quicker with the use of visual saliency . The most significant portions of the video can then be highlighted in the quality assessment by applying feature selection only to these noticeable areas . We can concentrate on the important areas rather than examining the complete video frame , which can be computationally costly . By selecting the most pertinent features from these categories , feature selection aids in producing more accurate and effective quality measurements . Salient locations frequently hold the most crucial data for evaluating quality . The quality assessment model performs better and is more resilient when redundant or less significant data is removed by feature selection in certain areas . By using visual saliency and feature selection , the process of evaluating video quality can be streamlined , resulting in more accurate findings that reflect the viewing experience and effective use of computational resources.

Table 5.1: Comparison of proposed methodology with state of the art Methods

Methodology	**MSE**	**PCC**	**KRCC**	**SROCC**
HVS-5M [22]	-	0.8422	-	0.8441
ReLaX-VQA [21]	-	0.8876	-	0.8468
MD-VQA [20]	**-**	0.839	**-**	0.814
XGC-VQA [24]	-	0.7805	0.7033	0.8245
Proposed	**-**	**0.997**	**0.955**	**0.995**

5.3 Affect of Feature Selection on Training Time

Feature selection decreases, the total exceution time for training different models.

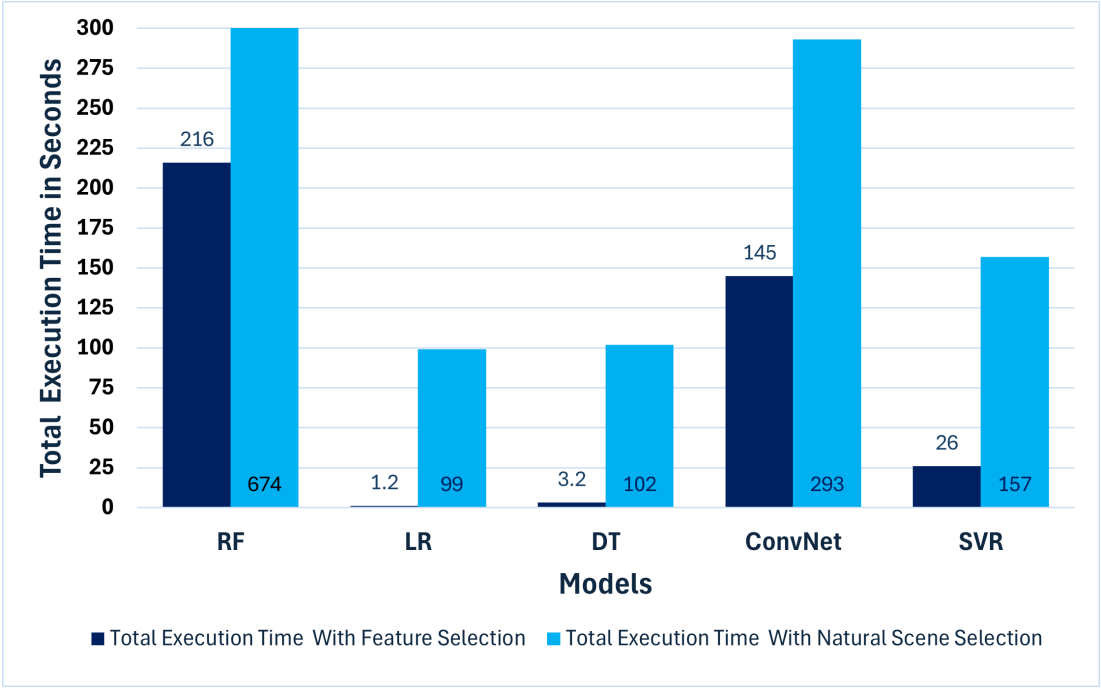


Figure 5.1: Performance comparison of models in terms of total execution time in seconds using FRIQUEE features

Figure 5.1 shows total execution time taken with natural scene selection and after performing feature selection. On the horizontal (x- axis), different regression models are defined, RF (Random Forest), LR (Linear Regression), DT (Decision Tree), ConvNet (Convolutional neural Network) and SVR (Support Vector Regression). On vertical (y-axis), total execution time in Seconds is defined.
It is evident from graph that attribute selection has decreased the execution time for RF (Random Forest) to 67.95 %, from (674 Seconds) to (216 Seconds). Execution time for LR (Linear Regression) has decreased to 98.7%, from (99 Seconds to 1.2 Seconds). For DT (Decision Tree), time has dropped to 96.8%, from (102 to 3.2 Seconds). ConvNet Similarly time has also decreased to 50.5%, from (293 to 145 Seconds). Similarly, for SVR, time has decreases about 83.43% (from 157 to 26 Seconds).

Figure 5.2 shows a graph between total execution time taken by different models, with and without feature selection. On the horizontal (x- axis), different regression models are defined, RF (Random Forest), LR (Linear Regression), DT (Decision Tree) and SVR (Support Vector Regression). On vertical (y- axis), total execution time in Seconds is defined. It can be observed in the graph that feature selection has decreased the execution time for RF (Random Forest) to 75.4%, from (75 Seconds) to (18.4 Seconds). Execution time for LR (Linear Regression) has decreased to 98.8%, from (15.6 Seconds to 0.18 Seconds). For DT (Decision Tree), time has dropped about 97.3%, from (0.43 to 16.1 Seconds). Similarly, for SVR about 62.3% decrease in time can be seen (26 to 9.8 Seconds).

Figure 5.3 shows a graph between total execution time taken by different models, with and without feature selection. On the horizontal (x- axis), different regression models are defined, RF (Random Forest), LR (Linear Regression), KNN (K Nearest Neighbor), ConvNet (Convolutional neural Network). On vertical (y- axis), total execution time in

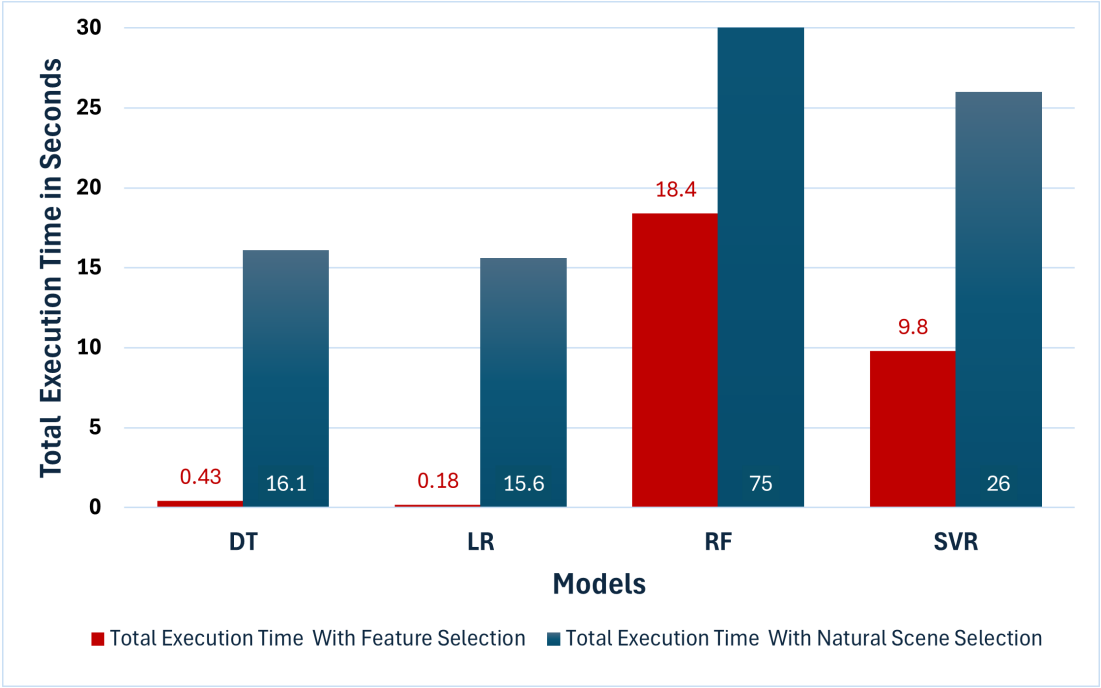


Figure 5.2: Performance comparison of models in terms of total execution time in seconds using GWH-GLBP features

Seconds is defined. it can be observed in the graph that feature selection has decreased the execution time for RF (Random Forest) to 50.6%, from (9.52 Seconds) to (4.7 Seconds). Execution time for LR (Linear Regression) has decreased to 99.3% (1.63 Seconds to 0.01 Seconds). For KNN (K Nearest Neighbor), time has dropped from (1.61 to 0.04 Seconds) about 97.5%. Similarly, for ConvNet, (17.9 to 8.17 Seconds) about 54.3% decrease in time can be seen.

Figure 5.4 shows a graph between total execution time taken by different models, with and without feature selection. On the horizontal (x- axis), different regression models are defined, RF (Random Forest), LR (Linear Regression), DT (Decision Tree), ConvNet (Convolutional neural Network) and SVR (Support Vector Regression). On vertical (y-axis), total execution time in Seconds is defined.

It can be observed in the graph that feature selection has decreased the execution time for RF (Random Forest) to 53.3%, from (44.6 Seconds) to (20.8 Seconds). Execution time for LR (Linear Regression) has decreased to 96.4%, from (5.7 Seconds to 0.2 Seconds). For DT (Decision Tree), time has dropped to 97%, from (6.7 to 0.2 Seconds). Similarly, for SVR (Support Vector Regression), 36.7% decrease in time can be seen (15.8 to 10 Seconds). For ConvNet, Execution time with feature selection has increased from 20 to 123 Seconds. This is because of tuning of hyperparameters in ConvNet. In this case, number of epochs for feature selection were increased. Otherwise in general, feature selection decreases the execution time.

Figure 5.5 shows a graph between total execution time taken by different models, with and without feature selection. On the horizontal (x- axis), different regression models are defined, RF (Random Forest), LR (Linear Regression), DT (Decision Tree), ConvNet

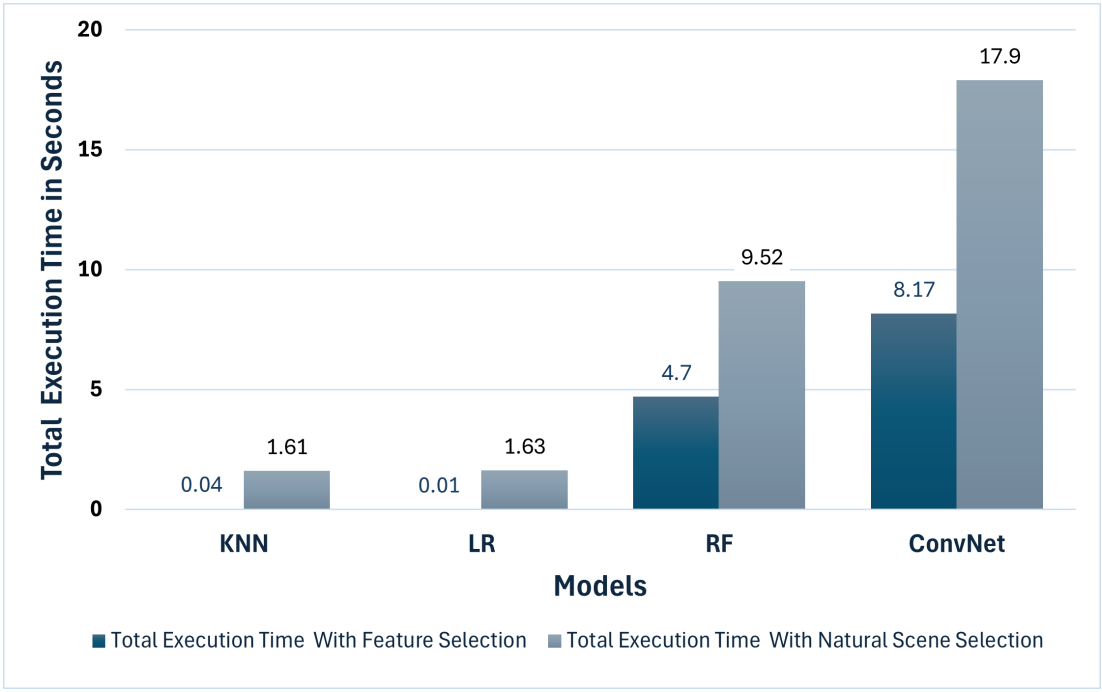


Figure 5.3: Performance comparison of models in terms of total execution time in seconds using OGIQA features

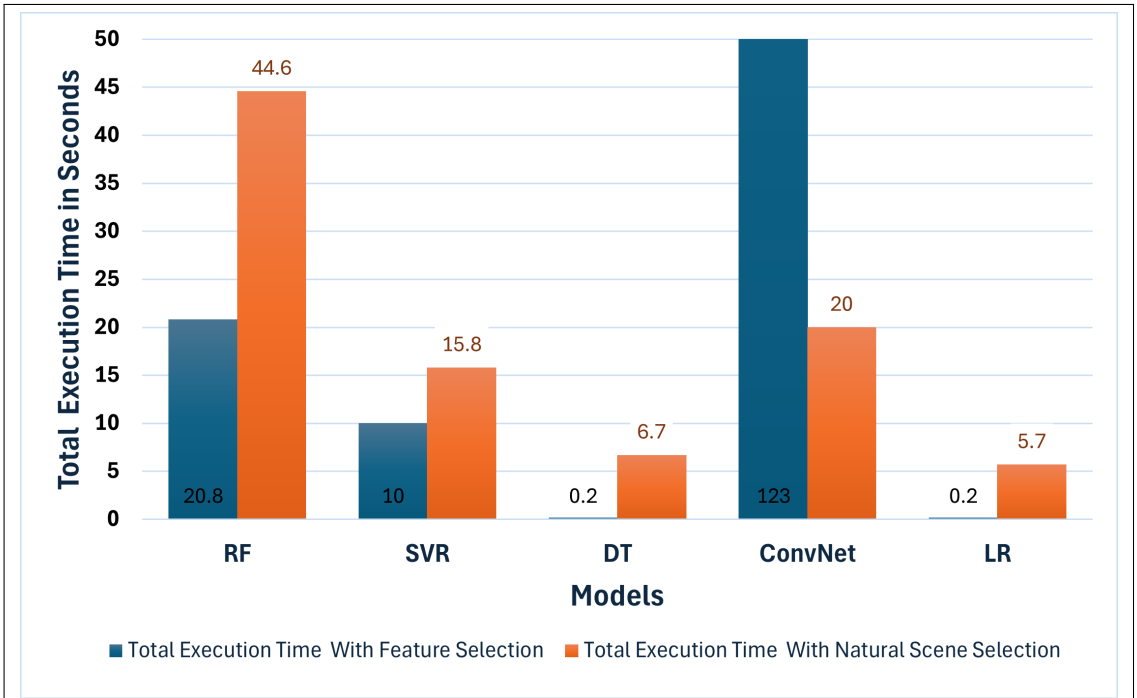


Figure 5.4: Performance comparison of models in terms of total execution time in seconds using BRISQUE features

(Convolutional neural Network), SVR (Support Vector Regression) and KNN (K Nearest Neighbor). On vertical (y- axis), total execution time in Seconds is defined.

It can be observed in the graph that feature selection has decreased the execution time for RF (Random Forest) to 71.2%, from (16.4 Seconds) to (4.71 Seconds). Execution time

for LR (Linear Regression) has decreased to 98.8%, from (2.6 Seconds to 0.03 Seconds). For DT (Decision Tree), time has dropped to 98%, from (3.08 to 0.06 Seconds). Execution time for KNN (K Nearest Neighbor) has decreased to 97.7%, from 2.8 to 0.07 Seconds. Similarly, for SVR (Support Vector Regression), a drop of 16.6% in execution time can be seen (12 to 10 Seconds). For ConvNet, Execution time with feature selection has increased from 14.1 to 96 Seconds. This is because of tuning of hyper parameters in ConvNet. In this case, number of epochs for feature selection were increased. Otherwise in general, feature selection decreases the execution time.

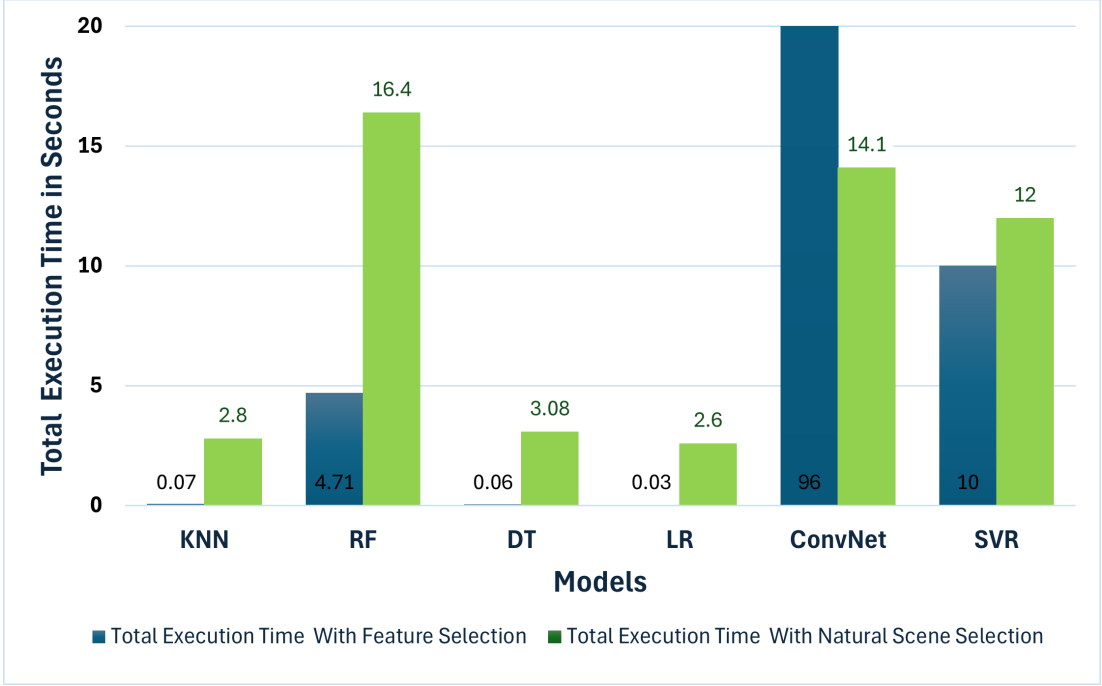


Figure 5.5: Performance comparison of models in terms of total execution time in seconds using CURVELET features

5.4 Future Recommendations

Visual Saliency has an important role to improve the video quality assessment . There is room for investigating other variables that might improve the outcomes while preserving computational efficiency . Choice of visual saliency techniques has significant impact on the outcomes . Therefore , the future research work can incorporate other visual saliency techniques to study their impact on the results . Subsequent research endeavors may concentrate on investigating the integration of visual saliency with other sophisticated procedures , crafting novel assessment measures , and implementing these approaches in diverse real-world contexts.

REFERENCES

[1] S. Singhal and M. Jena, "A study on weka tool for data preprocessing, classification and clustering," *International Journal of Innovative technology and exploring engineering (IJItee)*, vol. 2, no. 6, pp. 250–253, 2013. Cited on pp. xii and 22.

[2] Z. Sinno and A. Bovik, "Laboratory for Image and Video Engineering - The University of Texas at Austin — live.ece.utexas.edu." https://live.ece.utexas.edu/research/LIVEVQC/index.html, 2018. [Accessed 07-03-2024]. Cited on pp. xii and 26.

[3] I. F. Nizami, M. U. Rehman, A. Waqar, and M. Majid, "Impact of visual saliency on multi-distorted blind image quality assessment using deep neural architecture," *Multimedia Tools and Applications*, vol. 81, no. 18, pp. 25283–25300, 2022. Cited on p. 1.

[4] I. Bakurov, M. Buzzelli, R. Schettini, M. Castelli, and L. Vanneschi, "Full-reference image quality expression via genetic programming," *IEEE Transactions on Image Processing*, vol. 32, pp. 1458–1473, 2023. Cited on p. 2.

[5] C. Lee, J. Youn, and S. Woo, "Hybrid reduced-reference video quality assessment of streaming services over reliable transport," in *2023 14th International Conference on Information, Intelligence, Systems Applications (IISA)*, pp. 1–6, 2023. Cited on p. 2.

[6] H. Motamednia, P. Cheraaqee, A. Mansouri, and A. Mahmoudi-Aznaveh, "Quality assessment of screen content videos," in *2023 6th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pp. 1–7, 2023. Cited on p. 6.

[7] H. Su, Q. Liu, H. Yuan, Q. Cheng, and R. Hamzaoui, "Support vector regression-based reduced- reference perceptual quality model for compressed point clouds," *IEEE Transactions on Multimedia*, vol. 26, pp. 6238–6249, 2024. Cited on p. 6.

[8] Y. Mi, Y. Li, Y. Shu, and S. Liu, "Ze-fesg: A zero-shot feature extraction method based on semantic guidance for no-reference video quality assessment," in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3640–3644, 2024. Cited on p. 6.

[9] N.-W. Kwong, Y.-L. Chan, S.-H. Tsang, Z. Huang, and K.-M. Lam, "Deep learning approach for no-reference screen content video quality assessment," *IEEE Transactions on Broadcasting*, vol. 70, no. 2, pp. 555–569, 2024. Cited on p. 6.

[10] R. Tu, G. Jiang, M. Yu, T. Luo, Z. Peng, and F. Chen, "V-pcc projection based blind point cloud quality assessment for compression distortion," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 7, no. 2, pp. 462–473, 2023. Cited on p. 6.

[11] L. Zheng, J. Han, and Y. Xu, "A saliency map approach to optimize vmaf for video and image compression," in *2024 Data Compression Conference (DCC)*, pp. 293–301, 2024. Cited on p. 7.

[12] W. Wang and E. Dai, "Multi-view video quality enhancement method based on multi-scale fusion convolutional neural network and visual saliency," *IEEE Access*, vol. 12, pp. 33100–33108, 2024. Cited on p. 7.

[13] A. Rizal, A. Suharso, P. Abujabbar, and M. Munir, "Objective quality assessment of multi-resolution video based on h. 264/avc and h. 265/hevc encoding," in *Proceedings of the 7th Mathematics, Science, and Computer Science Education International Seminar, MSCEIS 2019, 12 October 2019, Bandung, West Java, Indonesia*, 2020. Cited on p. 7.

[14] J. Korhonen, "Two-level approach for no-reference consumer video quality assessment," *IEEE Transactions on Image Processing*, vol. 28, no. 12, pp. 5923–5938, 2019. Cited on p. 7.

[15] A. D. Andrushia and R. Thangarajan, "Center bias enhanced visual saliency detection method," in *2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN)*, pp. 1–4, IEEE, 2017. Cited on p. 7.

[16] J. Hu, J. Xiong, Y. Feng, and B. O. Onasanya, "Visual saliency detection based on visual center shift," in *2021 13th International Conference on Advanced Computational Intelligence (ICACI)*, pp. 227–232, IEEE, 2021. Cited on p. 7.

[17] W. Shen, M. Zhou, X. Liao, W. Jia, T. Xiang, B. Fang, and Z. Shang, "An end-to-end no-reference video quality assessment method with hierarchical spatiotemporal feature representation," *IEEE Transactions on Broadcasting*, vol. 68, no. 3, pp. 651–660, 2022. Cited on p. 7.

[18] Q. Kuang, X. Jin, Q. Zhao, and B. Zhou, "Deep multimodality learning for uav video aesthetic quality assessment," *IEEE Transactions on Multimedia*, vol. 22, no. 10, pp. 2623–2634, 2019. Cited on p. 7.

[19] Z. Tu, Y. Wang, N. Birkbeck, B. Adsumilli, and A. C. Bovik, "Ugc-vqa: Benchmarking blind video quality assessment for user generated content," *IEEE Transactions on Image Processing*, vol. 30, pp. 4449–4464, 2021. Cited on p. 8.

[20] Z. Zhang, W. Wu, W. Sun, D. Tu, W. Lu, X. Min, Y. Chen, and G. Zhai, "Md-vqa: Multi-dimensional quality assessment for ugc live videos," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1746–1755, 2023. Cited on pp. 8 and 51.

[21] X. Wang, A. Katsenou, and D. Bull, "Relax-vqa: Residual fragment and layer stack extraction for enhancing video quality assessment," *arXiv preprint arXiv:2407.11496*, 2024. Cited on pp. 8 and 51.

[22] A.-X. Zhang, Y.-G. Wang, W. Tang, L. Li, and S. Kwong, "Hvs revisited: A comprehensive video quality assessment framework," *arXiv preprint arXiv:2210.04158*, 2022. Cited on pp. 8 and 51.

[23] R. Rodrigues, L. Lévêque, J. Gutiérrez, H. Jebbari, M. Outtas, L. Zhang, A. Chetouani, S. Al-Juboori, M. Martini, and A. M. Pinheiro, "Objective quality assessment of medical images and videos: Review and challenges," *arXiv preprint arXiv:2212.07396*, 2022. Cited on p. 8.

[24] X. Huang, C. Li, A. Bentaleb, R. Zimmermann, and G. Zhai, "Xgc-vqa: A unified video quality assessment model for user, professionally, and occupationally-generated content," in *2023 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pp. 434–439, 2023. Cited on pp. 8 and 51.

[25] M. Zeeshan, M. Majid, I. F. Nizami, S. M. Anwar, I. U. Din, and M. K. Khan, "A newly developed ground truth dataset for visual saliency in videos," *IEEE Access*, vol. 6, pp. 20855–20867, 2018. Cited on p. 10.

[26] T. You and Y. Tang, "Visual saliency detection based on adaptive fusion of color and texture features," in *2017 3rd IEEE International Conference on Computer and Communications (ICCC)*, pp. 2034–2039, 2017. Cited on p. 10.

[27] P. Ye, Y. Wang, and Y. Xia, "Enhanced saliency prediction via orientation selectivity," in *2020 IEEE International Conference on Visual Communications and Image Processing (VCIP)*, pp. 91–95, 2020. Cited on p. 10.

[28] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 569–582, 2015. Cited on pp. 10 and 12.

[29] W.-P. Ma, W.-X. Li, J.-C. Sun, and P.-X. Cao, "Saliency detection via manifold ranking based on robust foreground," *International Journal of Automation and Computing*, vol. 18, pp. 73–84, 2021. Cited on p. 11.

[30] J. Yang, Y. Wang, G. Wang, and M. Li, "Salient object detection based on global multi-scale superpixel contrast," *IET Computer Vision*, vol. 11, no. 8, pp. 710–716, 2017. Cited on p. 11.

[31] J. Hao, J. Li, C. Pan, C. Huang, T. Xu, and H. Cheng, "Leveraging global and background contrast for salient object detection in hyperspectral images," in *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*, pp. 2186–2190, 2020. Cited on p. 12.

[32] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "ogiqa release." http://live.ece.utexas.edu/research/quality/og-iqa_release.zip,2015, 2015. [Accessed 05-03-2024]. Cited on p. 13.

[33] L. Liu, Y. Hua, Q. Zhao, H. Huang, and A. C. Bovik, "Blind image quality assessment by relative gradient statistics and adaboosting neural network," *Signal Processing: Image Communication*, vol. 40, pp. 1–15, 2016. Cited on p. 15.

[34] M. W. Aziz, I. Fareed Nizami, and M. Majid, "Blind quality assessment of super-resolution images using relative gradient statistics of salient and non-salient objects," in *2021 International Conference on Information Technology (ICIT)*, pp. 846–851, 2021. Cited on p. 15.

[35] Q. Li, W. Lin, and Y. Fang, "No-reference quality assessment for multiply-distorted images in gradient domain," *IEEE Signal Processing Letters*, vol. 23, no. 4, pp. 541–545, 2016. Cited on p. 16.

[36] D. Ghadiyaram and A. C. Bovik, "Perceptual quality prediction on authentically distorted images using a bag of features approach," *Journal of Vision*, vol. 17, pp. 32–32, 01 2017. Cited on p. 17.

[37] D. Ghadiyaram and A. C. Bovik, "Friquee software release." D.GhadiyaramandA.C.Bovik,"FRIQUEESoftwareRelease"URL:http://live.ece.utexas.edu/research/quality/FRIQUEE_release.zip,2016, 2016. [Accessed 05-03-2024]. Cited on p. 17.

[38] L. Liu, H. Dong, H. Huang, and A. C. Bovik, "No-reference image quality assessment in curvelet domain," *Signal Processing: Image Communication*, vol. 29, no. 4, pp. 494–505, 2014. Cited on p. 19.

[39] 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. Cited on p. 20.

[40] A. Mittal, A. K. Moorthy, and A. C. Bovik, "Brisque software release." A.Mittal,A.K.MoorthyandA.C.Bovik,"BRISQUESoftwareRelease",URL:http://live.ece.utexas.edu/research/quality/BRISQUE_release.zip,2011, 2011. [Accessed 05-03-2024]. Cited on p. 20.

[41] I. A. Siradjuddin, R. Septasurya, M. K. Sophan, N. Ifada, and A. Muntasa, "Feature selection with genetic algorithm for alcoholic detection using electroencephalogram," in *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, pp. 230–234, 2017. Cited on p. 21.

[42] F. Azuaje, I. Witten, and F. E, "Witten ih, frank e: Data mining: Practical machine learning tools and techniques," *Biomedical Engineering Online - BIOMED ENG ONLINE*, vol. 5, pp. 1–2, 01 2006. Cited on p. 21.

[43] P. Han, D.-B. Wang, and Q.-G. Zhao, "The research on chinese document clustering based on weka," in *2011 International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 1953–1957, 2011. Cited on p. 21.

[44] N. F. Sakina Rosdi, N. Shafiqah Ibrahim, I. H. Shamsudin, S. Mutalib, and S. Abdul-Rahman, "A provisional study of data mining classification algorithms in predicting credit card defaulters using weka tools," in *2023 IEEE 8th International Conference on Recent Advances and Innovations in Engineering (ICRAIE)*, pp. 1–5, 2023. Cited on p. 21.

[45] M. A. Hall and L. A. Smith, "Practical feature subset selection for machine learning," 1998. Cited on p. 21.

[46] Z. Sinno and A. C. Bovik, "Large-scale study of perceptual video quality," *IEEE Transactions on Image Processing*, vol. 28, no. 2, pp. 612–627, 2018. Cited on p. 26.

[47] Z. Sinno and A. C. Bovik, "Large scale subjective video quality study," in *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 276–280, IEEE, 2018. Cited on p. 26.

[48] Z. Bobbitt, "MSE vs. RMSE: Which metric should you use?." https://www.statology.org/mse-vs-rmse/, Sept. 2021. Accessed: 2024-7-24. Cited on p. 27.

[49] L. H. Alzubaidi, V. Malathy, A. H. A. Hussein, B. S. Z. Abood, and N. Tamilarasi, "Iris recognition using machine learning based pearson's correlation coefficient feature selection with hamming distance," in *2023 3rd International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, pp. 1–5, 2023. Cited on p. 28.

[50] J. D. Gibbons and S. Chakraborti, *Nonparametric statistical inference: revised and expanded*. CRC press, 2014. Cited on p. 28.

[51] T. Zhou, Z. Mei, X. Zhu, and Z. Huang, "Synchrony detection of epileptic eeg signals based on attention and pearson's correlation coefficient," in *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 531–535, 2020. Cited on p. 28.

[52] L. Zhu, "Selection of multi-level deep features via spearman rank correlation for synthetic aperture radar target recognition using decision fusion," *IEEE Access*, vol. 8, pp. 133914–133927, 2020. Cited on p. 28.

[53] E. Popovic, I. Zeger, M. Grgic, and S. Grgic, "Evaluation of color saturation and hue effects on image quality," in *2023 International Symposium ELMAR*, pp. 7–12, 2023. Cited on p. 29.

[54] H. Abdi, "The kendall rank correlation coefficient," *Encyclopedia of measurement and statistics*, vol. 2, pp. 508–510, 2007. Cited on p. 29.

[55] S. Kumar and I. Chong, "Correlation analysis to identify the effective data in machine learning: Prediction of depressive disorder and emotion states," *International Journal of Environmental Research and Public Health*, vol. 15, no. 12, 2018. Cited on p. 30.