International Journal of Computer Trends and Technology
ISSN: 2231–2803 / https://doi.org/10.14445/22312803/IJCTT-V71I9P105

Volume 71 Issue 9, 33-40, September 2023

Original Article

Assessment of Technical Information Quality using Machine Learning

Arvind Kumar Bhardwaj[1], Sandeep Rangineni[2], Divya Marupaka[3]

[1]*Information Technology, Capgemini, Texas, USA.*
[2]*Information Technology, Pluto TV, California, USA.*
[3]*Information Technology, Unikon IT, California, USA.*

Received: 16 July 2023 Revised: 28 August 2023 Accepted: 12 September 2023 Published: 30 September 2023

Abstract - *Even specialists sometimes do not comprehend the reasoning behind the choices made by the most advanced ML systems, making them opaque to end-users in high-stakes fields like medical diagnosis, financial decision-making, and others. Because of this, there has been a rise in attention paid to the problem of explaining ML, both in the academic world and in the fields where it is really useful. From a survey of explanatory theories, we isolate some characteristics. Metrics used for assessments are aimed at achieving the defined qualities of explainability. Developing a set of assessment measures that can be used across all available explanation approaches is impossible. Software's prevalence in consumer goods and services and its complexity have both been on the increase in recent years. As our reliance on software grows, so does the significance of monitoring, improving, and enhancing its quality. In order to monitor and manage different aspects of software systems, software metrics provide a quantifiable technique for doing so. The challenge of predicting software quality may be recast as one of categorization or concept learning within the framework of machine learning. In this study, we provide the groundwork for using machine learning techniques in big software companies for evaluating and forecasting product quality. We also provide evidence that machine learning techniques may be useful in this context. Some objective measures for evaluating image quality are hard and time-consuming to calculate because they rely on explicit modelling of the extremely non-linear nature of human perception. Even though ML-based techniques for visual quality evaluation have been shown to work in a number of studies, the general reliability of these paradigms remains unclear due to their susceptibility to overfitting. A thorough familiarity with the benefits and drawbacks that define learning machines is necessary before attempting to use ML to model perceptual systems. The best procedures are shown and exemplified in this work.*

Keywords - *Data analysis, Data preparation, Machine learning, Data collection, Visual quality, Software quality.*

1. Introduction

One of the primary aims of modern electronic entertainment gadgets is to provide the consumer with a fantastic experience. It is assumed that any technology used in transmitting and delivering digital media would either keep the original visual quality of the media unchanged or, ideally, improve it. Therefore, it is vital that multimedia transmission systems have checkpoints and, if required, restoration mechanisms for visual quality [1].

A positive user experience can only be ensured if the standards used to regulate visual quality are consistent with human perception. Perceptual processes underpinning the human visual system must be faithfully reproduced by systems tasked with evaluating. In order to automatically and objectively evaluate the quality of images and videos, several methods have been created and published. Typically, this is done by using a regression function to be fitted on objective truth data, such as ratings of perceived quality.

Most of these methods are complicated and time-consuming to compute because they increase dependability by explicitly modelling the HVS's extremely non-linear behaviour. Thus, most objective quality evaluation techniques are either inaccurate or too sophisticated for real-time settings [2]. We will first provide a comprehensive overview of current methods to achieve this goal before commenting and demonstrating how to implement an ML-based objective quality measure [4].

We examine some outstanding problems with ML-based VQA [1-7]. The last thoughts are presented. Second, visual quality evaluation using artificial intelligence. Theoretical models developed via years of ML study have proven useful in various fields, including computer vision, data mining, and bioinformatics. As a result, the VQA community has been more interested in the past decade in using ML technologies to modelling the HVS's fundamental perceptual processes [8-19].

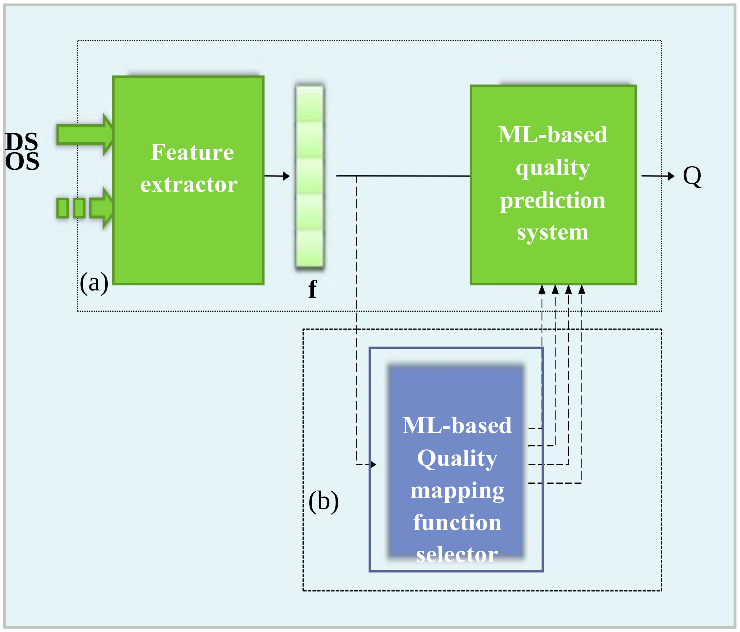

Fig. 1 ML-based image/video quality assessment system. The input (distorted) signal DS and, if available, the original signal (OS) are first represented in a low-dimensional feature space and then mapped into the quality space by means of a ML tool. Path (a) includes the two basic modules: 'feature extractor' and 'prediction system'; path (b) augments the framework with a module that is specifically designed

However, ML has its limits, which become apparent when used naively. Customers want to know the thinking behind the judgements made by ML models. Consequently, there is significant pressure from society and ethics to explain the workings of such ML systems.

When it comes to building trust and confidence in ML systems, explanations of ML outcomes are becoming more important, both for interpreting "black box" findings. From both an algorithmic and a visual analytics vantage point, many methods are presented for elucidating ML's workings [3]. This was before the period when Deep Learning was widely implemented. Existing studies have so far made preliminary efforts to establish methodologies to explain ability evaluations.

However, no standardized evaluation criteria for explanatory approaches make comparisons challenging. Note that the explanation is also domain-specific; there is no such thing as a universal explanation, and various kinds of explanations may serve different purposes. If you were turned down for a mortgage by your bank, for instance, you might want to know the major reasons why, but in a legal setting, you may need a more detailed explanation, including a list of all aspects.

This article provides an opinion on the metrics currently available to evaluate ML explanations and how to evaluate ML explanations in practice. The purpose of this work is to conduct a comprehensive literature analysis with the goal of classifying the current ways to evaluate [6].

We can determine its characteristics by looking at many contemporary definitions of explainability. Evaluation metrics are aimed at improving explaining ability based on the recognized qualities. Towards offering a comprehensive picture of the landscape of ML explanations, we also explore methods for visualizing ML explanations. Following this, part outlines the reasons for evaluating ML explanations and classifies the several ways ML explanations may be evaluated. The segment then goes on to discuss evaluations based on functionality. Section then examines evaluation metrics based on application, and part then reviews evaluation metrics based on human experience [5].

Before drawing any conclusions in the portion, we provide in-depth comments in which we point out areas of weakness and suggest new avenues for investigation. Case studies in ML often utilize a variety of words to describe the steps taken to shed light on ML's opaque workings to comprehend its deliberative procedures better. Words like "comprehensibility," "intelligibility," "transparency," and "understandability" come to mind. Other similar concepts include causality, which in Pearl means the connection between a cause and its result [7].

While causability does indeed relate to a human model, it measures how close an explanation comes to eliciting a certain degree of causal knowledge from a person. The term "causability" was chosen as a nod to the long-standing and widely accepted notion of "usability" in the field of software engineering. Many people use explainability and interpretability interchangeably, and these two concepts are used more often than others. Clarity and simplicity are two characteristics of interpretability. An unambiguous explanation is what we mean by "clarity," whereas "parsimony" refers to an explanation that is delivered in a brief and straightforward manner [8].

Lombrozo has shown that credible explanations tend to be brief and comprehensive. Another aspect of interpretability is broadness or the extent to which an explanation may be applied universally. Moreover, Reference claims that Fidelity has completeness and soundness. This explanation is complete and sound if it accurately depicts the ML model's dynamics. The idea of explainability and its associated features are shown in the figure. This study's taxonomy of assessment metrics is based on the concept of explainability, which is utilized throughout the rest of the paper [10].

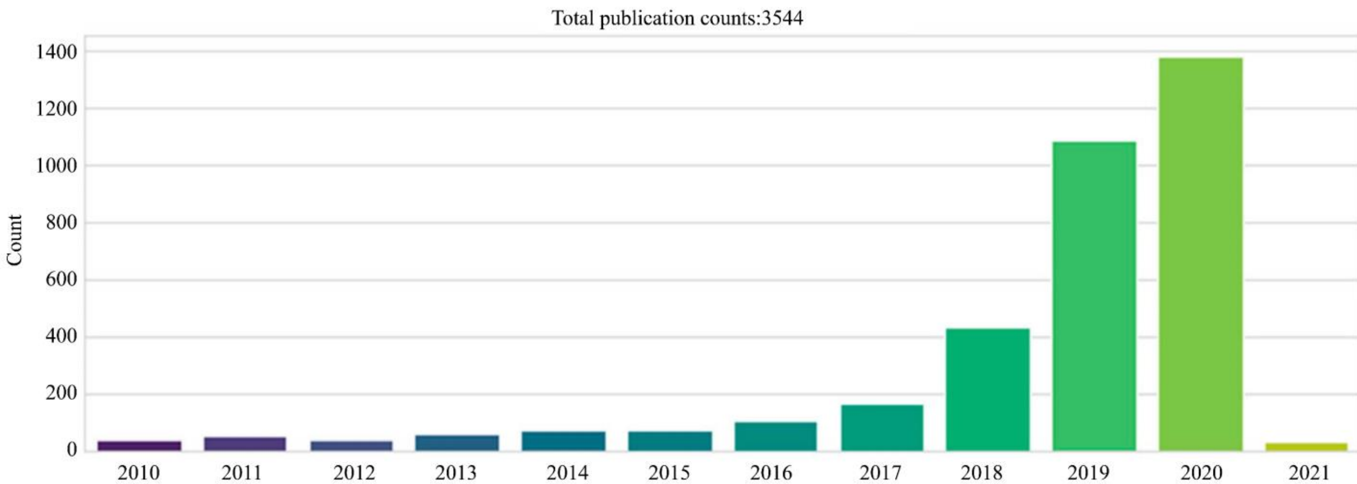

Fig. 2 The number of research publications on ML explanations (based on Scopus.com until December 2020)

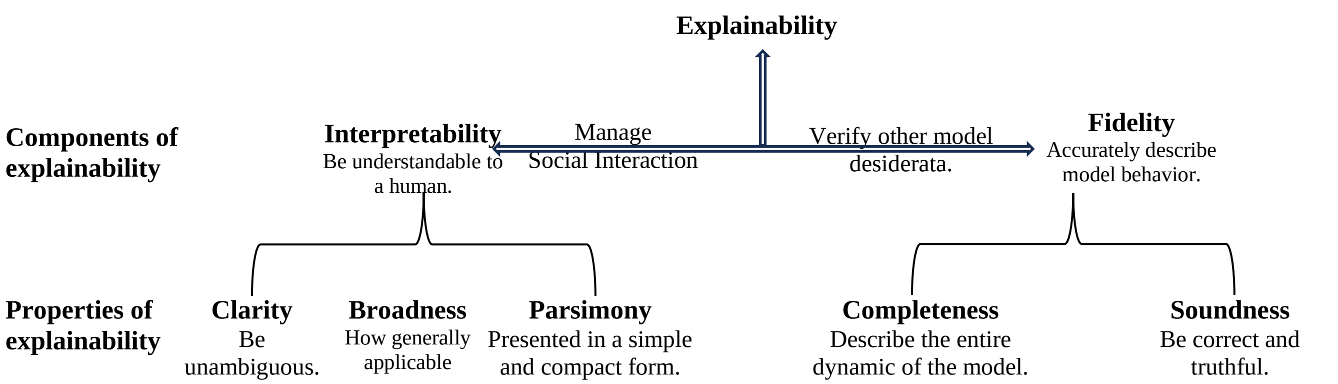

Fig. 3 Definition of machine learning (ML) explainability and related properties (adapted from Reference [27]).

1.1. Methods for Explaining Machine Learning

There has recently been much research on the capacity to explain machine learning results. Unlike other surveys, this one does not go further into the methods used for providing explanations. Instead, we put our attention where it belongs—on the evaluation's taxonomy of explanation methods [9].

1.2. Classification of Hypothesis Testing Methods

Various taxonomies have been suggested to classify explanations based on their origin, depth, breadth, and the types of models they can describe. Organized the explanations into classes according to the methods used, then linked those classes to their respective hierarchical placements. Methods, Materials, and Equipment Classification problems may be comprehended using various data components. Static and feature-based explanations are within the remit of local post hoc approaches. Methods aimed at visualizing neural networks.

This class of explanatory strategies is often used to explain neural networks and provides a visual picture of the network's intermediate representations/layers, moving from the static to the model to the global to the post hoc. It provides visualizations for inspecting a model's characteristics at a high level to grasp the model better. Since this is a worldwide phenomenon, it can only be explained in a static post hoc fashion. To better explain neural networks, for instance, Nguyen et al. presented multiple feature visualizations [11].

Saliency: Methods for determining the significance of features [static > modelling > global > post hoc > feature]: In this kind of explanatory approach, the connection between the input properties and the intended outcomes is investigated. This technique falls under the categories of static, global, and ad hoc justifications with accompanying visualizations. This class includes the Partial Dependence Plot, which displays the relatively little impact that one or two features have on the projected result. Using a single model, Hinton et al. (1996) successfully described a group of models, including neural networks. This class of explanations uses techniques like Variational AutoEncoder (VAE) and Generative Adversarial Network (GAN) to learn a high degree of interpretable characteristics from static data. To develop such interpretable representations, Chen et al. [40] presented a GAN-based technique. This kind of clarification is filed away in the features section after the data section [12].

Bounds Placed on Neural Network Designs: This class of justifications imposes constraints on the neural network design in order to make it understandable (static > model > global > direct). This may be categorized as a worldwide direct explanation. Zhang et al. suggested transforming a standard CNN into an explainable CNN by permitting unambiguous knowledge representations (such as a particular object portion) in the CNN's high cover layers [28]. Both the business and its employees may gain from the ML explanation. It may assist a business in maintaining legal compliance, gaining the confidence of its consumers, and strengthening its own internal controls. People gain as well because they are better educated, have better results, and are able to participate actively in making decisions that affect their lives. After discussing these potential advantages and disadvantages, references outlined six distinct justifications [13].

The Rationale for the Results: Explanations of this category focus on the personal and societal ramifications of ML system usage and decision-making. Such an explanation may help people feel more in charge of their role in decision-making aided by machine learning [15]. A person's involvement in the decision-making process and the potential results of the choice may be properly evaluated if he or she is aware of the potential repercussions of the decision. Using a single model, Hinton et al. (1996) successfully described a group of models, including neural networks. This class of explanations uses techniques like Variational Autoencoder (VAE) and Generative Adversarial Network (GAN) to learn a high degree of interpretable characteristics from static data. To develop such interpretable representations, Chen et al. [40] presented a GAN-based technique. This kind of clarification is filed away in the features section after the data section [19].

The Responsibility for the Results: Both the business and its employees may gain from the ML explanation. It may assist a business in maintaining legal compliance, gaining the confidence of its consumers, and strengthening its own internal controls. People gain as well because they are better educated, have better results, and are able to participate actively in making decisions that affect their lives. After discussing these potential advantages and disadvantages, references then outlined six distinct sorts of justifications [20].

Explanation of Reasons: A "why" explanation for ML decisions explains the thought process behind a call and is written with non-experts in mind. In cases when the ML conclusion was not what the user anticipated, this form of explanation might help them determine whether the decision was faulty. However, if this is the case, the explanation helps them to make fair justifications for why they hold this view.

Justification of Fault: This sort of justification addresses the "who" questions of an ML system's creation and administration. It also aids in the tracking of responsibility defining the data. This justification style emphasizes the data's nature and role in making a choice and the data's nature and role in training and testing the ML model. Users may benefit from this style of explanation because it clarifies the role data plays in informing choices and reasoning for fairness. It is important to know whether or not you have been treated fairly [21].

Performance and Safety Concerns: This kind of justification is essential for boosting people's trust in an AI system. If one is shown how prejudice and bias are avoided in decision-making, it might increase their level of confidence in the system. Clarification on performance and safety. Explanations of this category focus on how an ML system's accuracy, dependability, security, and robustness were improved throughout its whole, from the design phase to implementation. The rationale for the result and explanations of this category focus on the personal and societal ramifications of ML system usage and decision-making. Such an explanation may help people feel more in charge of their role in decision-making aided by machine learning. A person's involvement in the decision-making process and the potential results of the choice may be properly evaluated if he or she is aware of the potential repercussions of the decision. Among them, the ethical standards of artificial intelligence distinct from explainability include accountability and fairness [22].

Effect Explanations: However, employing an ML system has consequences, which must be explained in terms of the effect. Therefore, ML's explaining skills are strongly tied to the ability to provide explanations for things like rationality, data, security and performance. Arya et al. classifications of explanatory strategies allow us to link various explanation varieties with specific strategies. The table below illustrates some common relationships between different kinds of explanation methods and the various explanation types presented in Reference [23].

It is possible that certain methods of explanation might be used for various kinds of justifications. Imprecision and judgements in statistics, since both the model and its justifications are built from data, machine learning introduces uncertainty into both. However, the fundamental purpose of explanation is to provide light on the underlying causal structure, but most statistical learning processes show correlation patterns between characteristics instead. The dependency on features compounds attribution and extrapolation errors. Misleading explanations may arise through extrapolation and associated characteristics.

2. Review of Literature

High-Level Outline of the Machine Learning-Based Quality Assessment Model

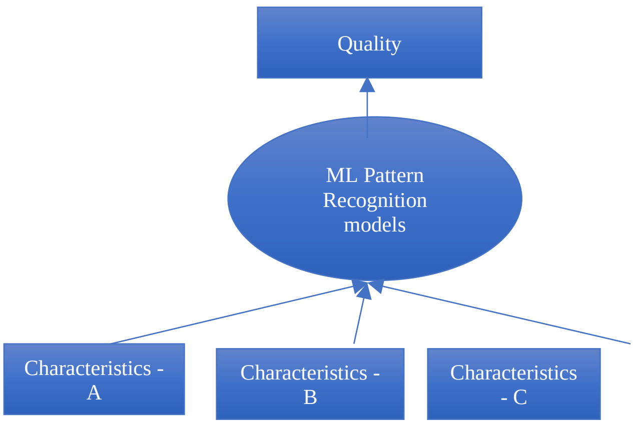

Fig. 4 Machine Learning-Based quality assessment model

We say that the attribution technique fulfils sensitivity- n. In this approach, the sensitivity-n is used to evaluate the quality of an explanation by determining whether or not the models meet the condition for all n. However, it is unclear how to use this criterion to evaluate the explanatory power of the various approaches. In ML, explanations often make use of perturbations. The suggested sensitivity metric would be used to evaluate how much small changes to the test point might alter the explanation. A smoothing of explanations was also suggested by Yeh et al. to increase sensitivity. Removing a single feature from the input and observing the resultant impact on the model's performance is a typical method for estimating the relevance of features [24].

3. Research, in Addition to Methodology

In a top-down fashion, as shown in the figure, this model would be heavily influenced by the measurement information model.

3.1. Comprehensive Review of an ML-Based Model for Evaluating High-Order Quality Attributes

The State of Quality Control in Software: As we rely more and more on software, the factors contributing to its quality have become more crucial. Like many other characteristics, quality may be significantly enhanced by careful definition and continual measurement.

Quality is one of the most often used concepts, yet it is also one of the most vague and misinterpreted. A federal court famously stated of obscenity, "I know it when I see it," and many people feel the same way about excellence [25].

Table 2. Characteristics and Sub characteristics of the internal/external quality model.

Characteristics	Sub Characteristics
Functionality	suitability
	accuracy
	interoperability
	security
	functionality compliance
Reliability	maturity
	fault tolerance
	recoverability
	reliability compliance
Usability	understandability
	learnability
	operability
	attractiveness
	usability compliance
Efficiency	time behavior
	resource efficiency utilization
	efficiency compliance
Maintainability	analyzability
	changeability
	stability
	testability
	maintainability compliance
Portability	adaptability
	installability
	co-existence
	replaceability
	portability compliance

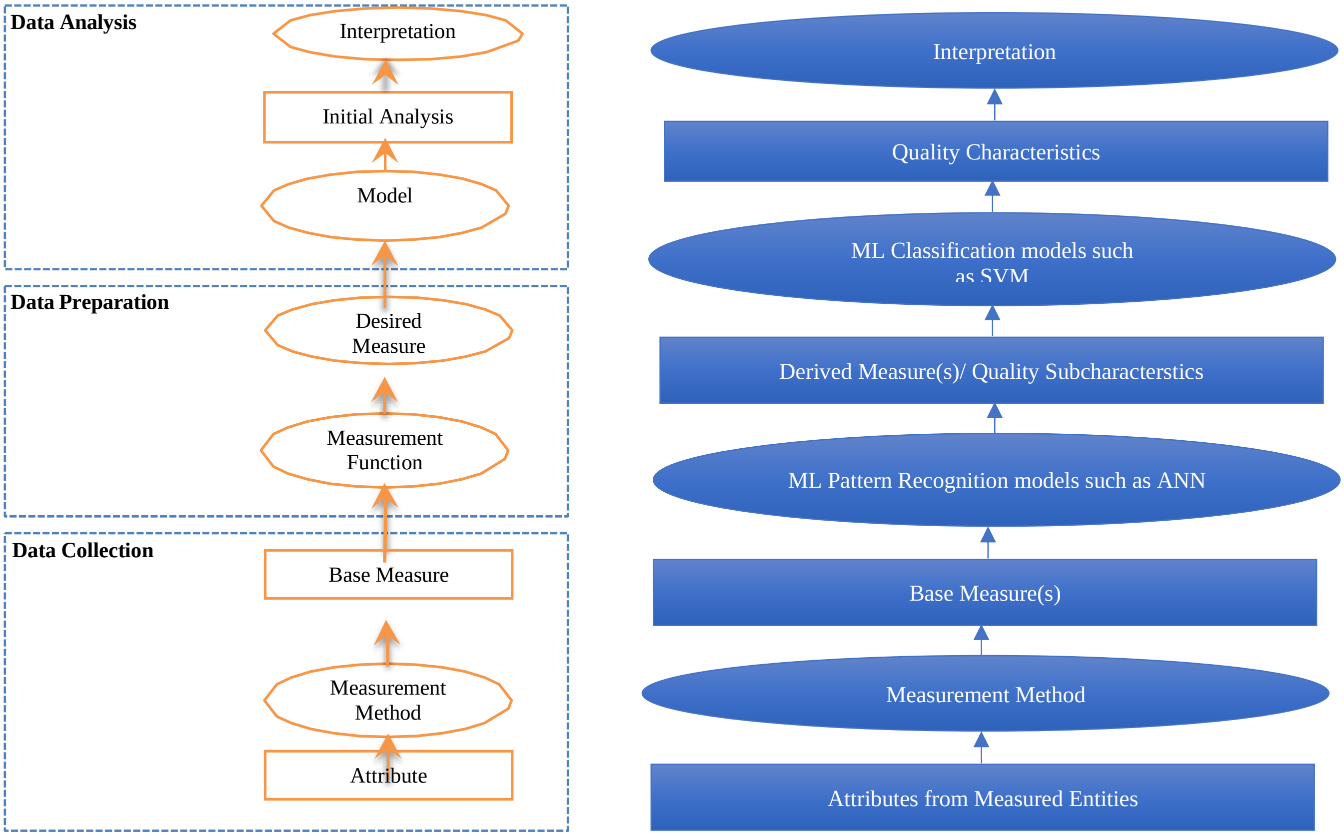

Fig. 5 Machine learning measurement information model

Table 3. Measure and Characteristics of the internal/external quality model

Measure	Characteristics
Quality in Use	Effectiveness
	Productivity
	Safety
	Satisfaction

Quality measurements, features, and sub-characteristics may be given varying weights to tailor the quality model to individual requirements based on the nature of the product, the context in which it will be used, or both metrics and quantification for accounts that rely on attribution [26].

The majority of the current ML literature is devoted to attribution-based explanations. As a result, several methods are provided with which to evaluate the accuracy of such accounts. The monotonicity and non-sensitivity of an explanation for attributing a trait are two such measures.

When the effective complexity of a system is minimal, it is easier to disregard some of the aspects despite their potential impact (lower cognitive salience) (non-sensitivity). Simple and broad characterize explanations with minimal effective complexity [27].

4. Conclusion

"Quality" is a ubiquitous but vague concept when applied to software. Although everyone has a sense of quality, what constitutes quality varies greatly depending on the context in which a product is utilized and the expectations of its end users. A large and well-established software development company will routinely collect and monitor software metrics as part of its day-to-day activity across all projects.

This work discusses the pros and cons of employing machine learning to evaluate visual quality. By surveying the various methods already in use, we were able to summarise the shared features of ML-based VQA, demonstrating how they can achieve results on par with or even superior to those of conventional approaches while also calling attention to the problems that can undermine their trustworthiness. Overall, it seems that ML may be quite useful in VQA. However, further study is required to perfect a few key areas.

To avoid the curse of dimensionality, we must first create and implement perception-oriented feature selection algorithms, and then we must construct rigorous validation processes to evaluate these systems, including defined phases like model selection and fair assessment of the generalization error. Last but not least, the availability of training data is a key challenge in building trustworthy ML-based VQA; hence, it is important to construct either re-alignment techniques or subjective approaches that enable the gathering of similar quality ratings across various trials. We defined explainability and then specified the requirements of ML explainability that need to be met for an explanation to be considered satisfactory.

We then established links between the characteristics of explainability and types of ML explanatory strategies. However, there are no established standards for evaluating human-centered designs of experiments or quantifying subjective outcomes objectively. In the end, we determined that examining ML justifications requires the collaboration of several academic fields.

References

[1] Fu-Huei Lin, and Russell M. Mersereau, "Rate–Quality Tradeoff MPEG Video Encoder," *Signal Processing: Image Communication*, vol. 14, no. 4, pp. 297–309, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[2] P. Gastaldo, S. Rovetta, and R. Zunino, "Objective Quality Assessment of MPEG-2 Video Streams by Using CBP Neural Networks," *IEEE Transactions on Neural Networks*, vol. 13, no. 4, pp. 939–947, 2002. [CrossRef] [Google Scholar] [Publisher Link]

[3] Susu Yao et al., "Video Quality Assessment Using Neural Network Based on Multi-Feature Extraction," *Visual Communications and Image Processing 2003*, vol. 5150, 2003. [CrossRef] [Google Scholar] [Publisher Link]

[4] P. Le Callet, C. Viard-Gaudin, and D. Barba, "A Convolutional Neural Network Approach for Objective Video Quality Assessment," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1316–1327, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[5] S. Kanumuri et al., "Modeling Packet-Loss Visibility in MPEG-2 Video," *IEEE Transactions on Multimedia*, vol. 8, no. 2, pp. 341–355, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[6] Hasnaa El Khattabi, Ahmed Tamtaoui, and Driss Aboutajdine, "Video Quality Assessment Measure with a Neural Network," *International Journal of Computer and Information Technology*, vol. 4, no. 3, pp. 167–171, 2010. [Google Scholar]

[7] Manish Narwaria, Weisi Lin, and Anmin Liu, "Low-Complexity Video Quality Assessment Using Temporal Quality Variations," *IEEE Transactions on Multimedia*, vol. 14, no. 3, pp. 525–535, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[8] Nicolas Staelens et al., "Constructing a No-Reference H.264/AVC Bitstream-Based Video Quality Metric Using Genetic Programming-Based Symbolic Regression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 8, pp. 1322–1333, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[9] Paolo Gastaldo, and Rodolfo Zunino, "Neural Networks for the No-Reference Assessment of Perceived Quality," *SPIE Journal of Electronic Imaging*, vol. 14, no. 3, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[10] S. Suresh, R. Venkatesh Babu, and H.J. Kim, "No-Reference Image Quality Assessment Using Modified Extreme Learning Machine Classifier," *Applied Soft Computing*, vol. 9, no. 2, pp. 541–552, 2009. [CrossRef] [Google Scholar] [Publisher Link]

[11] Manish Narwaria, and Weisi Lin, "Objective Image Quality Assessment Based on Support Vector Regression," *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[12] Judith A. Redi et al., "Color Distribution Information for the Reduced-Reference Assessment of Perceived Image Quality," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 12, pp. 1757–1769, 2010. [CrossRef] [Google Scholar] [Publisher Link]

[13] Hantao Liu et al., "An Efficient Neural-Network Based No-Reference Approach to an Overall Quality Metric for JPEG and JPEG2000 Compressed Images," *SPIE Journal of Electronic Imaging*, vol. 20, no. 4, pp. 1–15, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[14] Chaofeng Li, Alan Conrad Bovik, and Xiaojun Wu, "Blind Image Quality Assessment Using a General Regression Neural Network," *IEEE Transactions on Neural Networks*, vol. 22, no. 5, pp. 793–799, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[15] Anush Krishna Moorthy, and Alan Conrad Bovik, "Blind Image Quality Assessment: From Natural Scene Statistics to Perceptual Quality," *IEEE Transactions on Image Processing*, vol. 20, no. 12, pp. 3350–3364, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[16] Huixuan Tang, Neel Joshi, and Ashish Kapoor, "Learning a Blind Measure of Perceptual Image Quality," *IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs*, pp. 305-312, 2011. [CrossRef] [Google Scholar] [Publisher Link]

[17] Sergio Decherchi et al., "Circular-ELM for the Reduced-Reference Assessment of Perceived Image Quality," *Neurocomputing*, vol. 102, pp. 78–89, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[18] Christophe Charrier, Olivier Lézoray, and Gilles Lebrun, "Machine Learning to Design Full-Reference Image Quality Assessment Algorithm," *Signal Processing: Image Communication*, vol. 27, no. 3, pp. 209–219, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[19] Manish Narwaria, Weisi Lin, and A. Enis Cetin, "Scalable Image Quality Assessment with 2D Mel-Cepstrum and Machine Learning Approach," *Pattern Recognition*, vol. 45, no. 1, pp. 299–313, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[20] Manish Narwaria, and Weisi Lin, "SVD-Based Quality Metric for Image and Video Using Machine Learning," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 347–364, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[21] Thiede Thilo et al., "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *Audio Engineering Society*, vol. 48, no. 1/2, pp. 3–29, 2000. [Google Scholar] [Publisher Link]

[22] Nicu Sebe et al., *Machine Learning in Computer Vision*, Springer, 2005. [Google Scholar] [Publisher Link]

[23] Giovanni Da San Martino, and Alessandro Sperduti, "Mining Structured Data," *IEEE Computational Intelligence Magazine*, vol. 5, no. 1, pp. 42–49, 2010. [CrossRef] [Google Scholar] [Publisher Link]
[24] J.C. Rajapakse, Y-Q. Zhang, and G.B. Fogel, "Computational Intelligence Approaches in Computational Biology and Bioinformatics," *IEEE/ACM Transactions on Computational Biology and Bionformatics*, vol. 4, 2007. [Google Scholar]
[25] Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006. [Google Scholar] [Publisher Link]
[26] Isabelle Guyon, and Andre Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–11822, 2003. [Google Scholar] [Publisher Link]
[27] Aniek F. Markus, Jan A. Kors, and Peter R. Rijnbeek, "The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies," *Journal of Biomedical Informatics*, vol. 113, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[28] Vivek Basavegowda Ramu, "Performance Testing using Machine Learning," *SSRG International Journal of Computer Science and Engineering*, vol. 10, no. 6, pp. 36-42, 2023. [CrossRef] [Publisher Link]
[29] Vijay Arya et al., "One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques," *arXiv*, 2019. [CrossRef] [Google Scholar] [Publisher Link]