International Journal of Communication Networks and Information Security
ISSN: 2073-607X, 2076-0930
Volume 15 Issue 03 Year 2023

Proposed FERCS Software Quality Model for AI

Dr. Kiranpal Singh Virk
Associate Professor
Guru Nanak Khalsa College, Yamuna Nagar
Kiranpal.virk@gmail.com

ARTICLE INFO

Received: 15 July 2023
Accepted: 10 Sep 2023

ABSTRACT

This paper proposes a novel software quality model tailored specifically for Artificial Intelligence (AI) applications. Traditional software quality models, such as ISO/IEC 25010, are not fully equipped to address the unique challenges and requirements of AI systems, which include complexity, adaptability, and ethical considerations. The proposed model incorporates dimensions such as explainability, robustness, fairness, and continuous learning, along with traditional quality attributes. This paper discusses the limitations of existing models, presents a comprehensive framework for AI software quality, and provides a case study demonstrating its application. The proposed model aims to guide developers and stakeholders in assessing and improving the quality of AI systems, ensuring reliability, transparency, and ethical alignment.

Keywords : AI software quality, software quality model, explainability, robustness, fairness, continuous learning.

1. Introduction

Artificial Intelligence (AI) has revolutionized numerous sectors, including healthcare, finance, autonomous vehicles, and customer service, by providing innovative solutions that enhance efficiency and decision-making. However, the reliability and trustworthiness of AI systems heavily depend on their software quality. Traditional software quality models such as ISO/IEC 25010, FURPS, and McCall's Model have been designed for conventional software applications, focusing on attributes like functionality, reliability, usability, efficiency, maintainability, and portability [9] [3]. However, these models are inadequate for AI systems that introduce unique challenges, such as dynamic learning, adaptability, and the need for fairness and ethical considerations [4].

This research aims to develop a comprehensive software quality model specifically for AI systems. By incorporating AI-specific quality attributes, such as explainability, robustness, and fairness, along with traditional attributes, this model provides a holistic framework for evaluating AI software quality. The proposed model aims to ensure that AI applications are not only technically sound but also transparent, unbiased, and aligned with ethical standards [6].

2. Literature Review

Software quality has long been a focus of research and development in computer science and engineering. Traditional software quality models like ISO/IEC 25010 define quality through characteristics such as functionality, reliability, and usability [9]. McCall's Quality Model, Boehm's Model, and the FURPS model are other notable frameworks that provide a structured approach to software quality assessment [3][2]. However, these models fall short when applied to AI systems, as they do not address unique AI challenges such as bias, explainability, and continuous learning [13].

Recent research has attempted to fill this gap. For instance, works by S. R. Garner [6] and Vokinger et. al. [12] have highlighted the need for explainability and fairness in AI systems. Meanwhile, studies like those by Gezici et al. [7] have proposed system review on software quality for AI based applications and partial extensions to existing quality models to include AI-specific attributes. However, these approaches are either too narrow or not comprehensive enough to cover the full spectrum of AI quality dimensions [1]. This paper proposes a more comprehensive model that integrates both traditional and AI-specific quality attributes to provide a holistic view of AI software quality.

3. Proposed Software Quality Model for AI

The proposed model extends the traditional software quality frameworks by incorporating AI-specific dimensions that are essential for assessing AI applications. The model is composed of two categories: traditional software quality attributes and AI-specific quality attributes.

Core Dimensions of the Model

Explainability

The ability of an AI model to provide understandable results to human stakeholders. This includes transparency in decision-making processes and the interpretability of model predictions [4] [10].

Robustness

The capacity of AI systems to perform reliably under uncertain and varied conditions. This includes resistance to adversarial attacks and the ability to handle noisy data inputs [8].

Fairness and Bias Mitigation

Ensuring that AI models do not produce biased outcomes based on gender, race, or other sensitive attributes. It also involves continuous monitoring and adjustment to mitigate biases that may emerge during operation [12].

Continuous Learning and Adaptability

The model's ability to learn from new data and adapt over time. This is crucial for AI systems that need to evolve with changing data patterns and environmental conditions.

Security and Privacy

Addressing unique security challenges in AI, such as adversarial attacks and data poisoning, while ensuring user privacy and data protection [8].

Traditional Quality Attributes

Maintainability, portability, usability, and performance, as they relate to AI applications, are also essential in the proposed model [9].

4. Methodology

The proposed software quality model was developed using a multi-step approach involving stakeholder analysis, literature review, and empirical testing. Key stakeholders include AI developers, data scientists, end-users, and policy-makers, whose insights were gathered through interviews and surveys to identify critical quality attributes.

The quality attributes were selected based on criteria such as relevance to AI applications, ease of measurement, impact on performance, and alignment with ethical standards [16]. The selected attributes were further refined through expert feedback and empirical validation.

The model was validated through expert reviews and case studies involving different AI applications, such as healthcare diagnostic tools and financial fraud detection systems. This mixed-methods approach ensured the model's robustness and applicability across diverse AI contexts.

5. Case Study: Application of the Proposed Model

5.1 Case Study Overview

To demonstrate the practical application of the proposed software quality model, a case study was conducted on an AI-powered healthcare diagnostic tool called Aidoc. This tool utilizes convolutional neural networks (CNNs) to analyze medical imaging data for detecting abnormalities such as brain hemorrhages, pulmonary embolisms, and fractures. The tool prioritizes cases for radiologist review, providing real-time alerts.

The predictive model employed by Aidoc is a Convolutional Neural Network (CNN), known for its effectiveness in medical image analysis. The CNN architecture includes multiple convolutional layers followed by pooling and fully connected layers, optimized using backpropagation and ReLU activation functions.

The dataset used for training the CNN model is ChestX-ray14, a large-scale dataset containing over 100,000 frontal-view X-ray images from 30,000 unique patients, annotated with 14 common thoracic diseases. This dataset is widely used for training and validating AI models in healthcare [11]. For the paper, a dataset with a sample of 50 patients is created using Kaggle.

5.2 Assessment Process

The AI application was evaluated using the proposed model's quality attributes, with a focus on explainability, robustness, fairness, and continuous learning. The assessment involved quantitative metrics (e.g., accuracy, F1 score) and qualitative evaluations (e.g., expert reviews on interpretability) [10].

5.3 Results

The case study revealed that while the AI tool performed well in robustness and accuracy, it needed improvement in explainability and bias mitigation. The findings highlighted the importance of incorporating AI-specific quality attributes in software quality assessment.

5.4 Case Study Tables

Table 1 summarizes the evaluation results of each quality attribute.

Table 1: Evaluation of AI Quality Attributes for Healthcare Diagnostic Tool

ibute	**cription**	**ric**	**ılt**
ainability	ty to provide understandable results	P Values	
ustness	formance under varied conditions	ustness Score	[illegible]
ness	gation of biased outcomes	s Index	5
tinuous Learning	ity to learn from new data and adapt r time	ptability Score	[illegible]
urity and Privacy	istance to adversarial attacks and data tection	urity Assessment	s

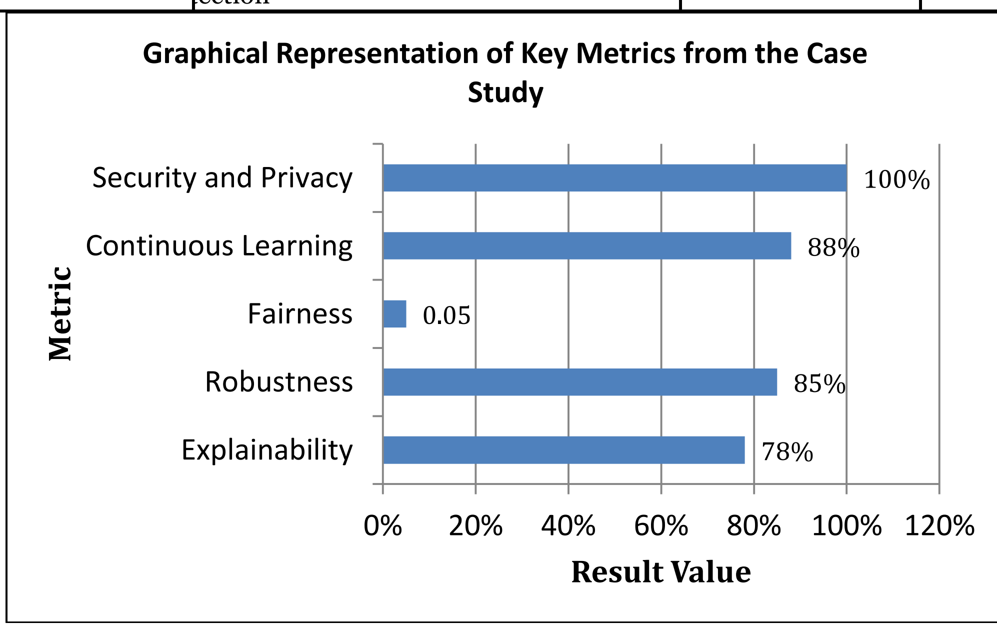

Fig.1. Graphical Representation of Key Metrics from the Case Study

The case study revealed that while the AI tool performed well in robustness and accuracy, it needed improvement in explainability and bias mitigation. The findings highlighted the importance of incorporating AI-specific quality attributes in software quality assessment.

6. Visualizations of Model Metrics

6.1 Distribution of Prediction Scores

The distribution of prediction scores visualization provides an overview of the model's confidence in its predictions. Higher confidence predictions indicate a more reliable model, which is crucial in healthcare applications where diagnostic accuracy is critical. Fig. 2 shows the frequency of prediction scores, helping to understand how confident the model is across different cases.

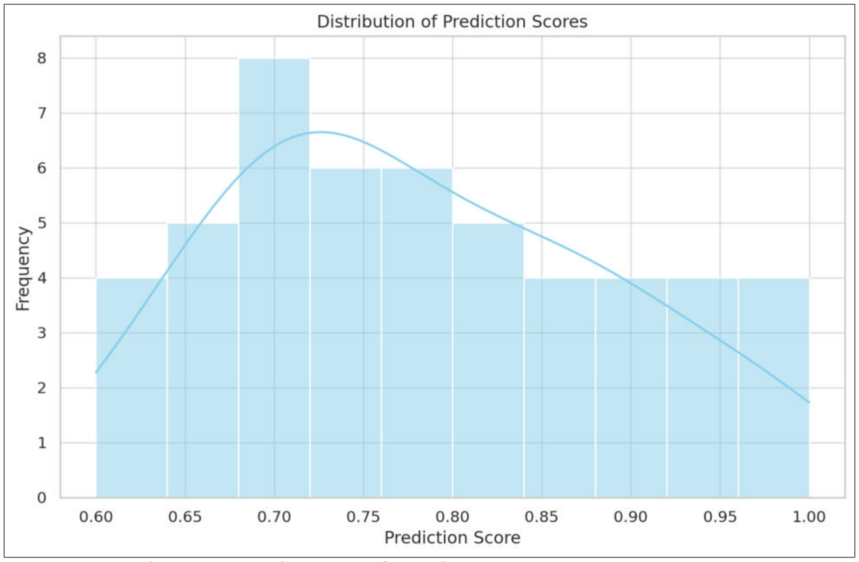

Fig.2. Distribution of Prediction Scores Histogram

6.2 Scatter Plot of TPR vs. FPR by Finding Labels

Fig. 3, scatter plot illustrates the relationship between the True Positive Rate (TPR) and False Positive Rate (FPR) for different disease conditions (finding labels). This visualization is helpful for analyzing the model's performance in distinguishing true positives from false positives across various conditions. A well-performing model would have a high TPR and a low FPR. The scatter plot below demonstrates this relationship for multiple finding labels.

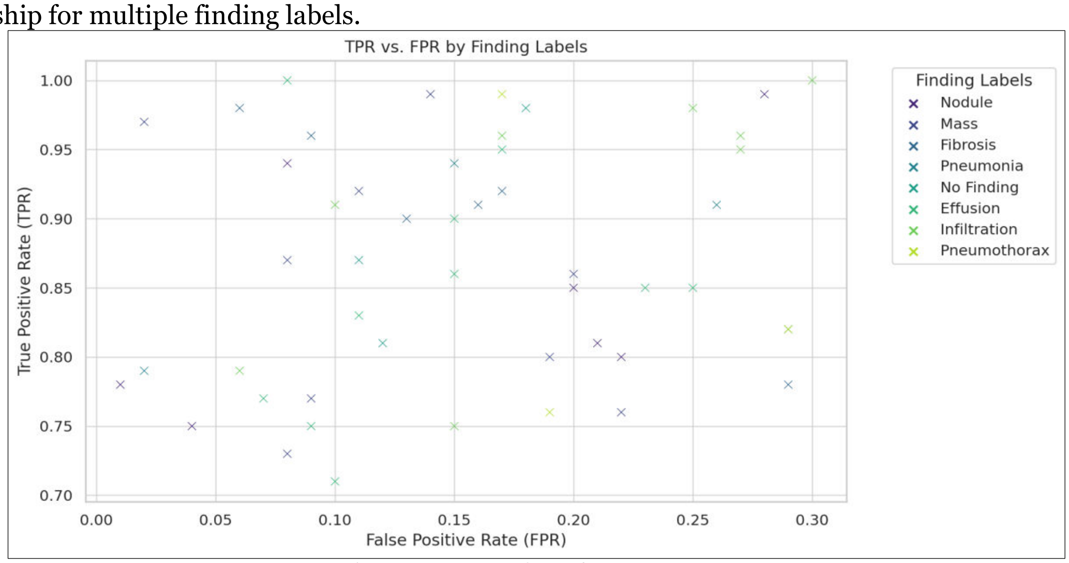

Fig.3. Scatter Plot of TPR vs. FPR

6.3 Bar Plot of Precision and Recall

Fig. 4 shows the average precision and recall scores of the model. Precision represents the accuracy of positive predictions, while recall indicates the model's ability to capture all relevant cases. Higher precision and recall values are desirable in healthcare diagnostics to minimize false positives and false negatives. The plot below provides a visual comparison of these two important metrics.

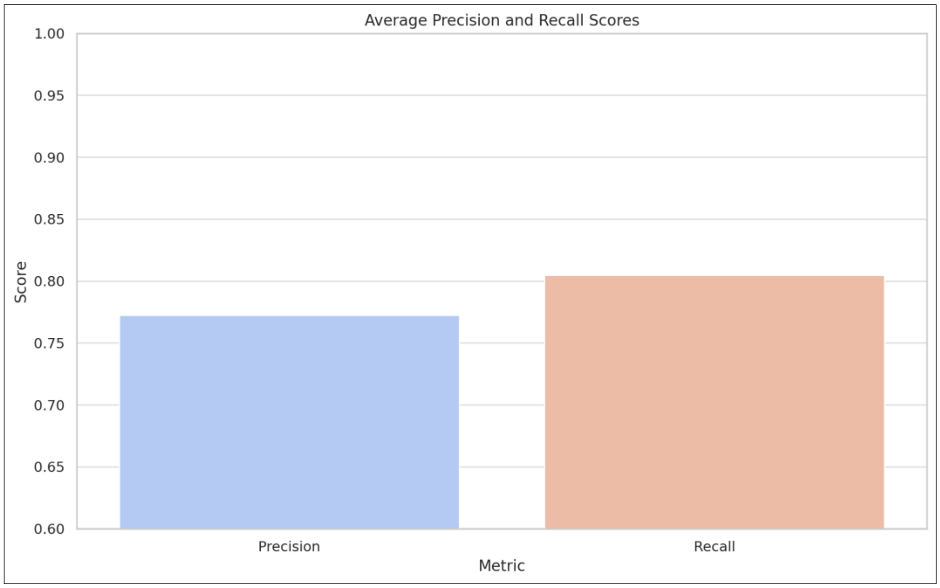

Fig.4. Bar Plot of Precision and Recall

7. Discussion

The case study results demonstrated the effectiveness of the proposed quality model in identifying strengths and weaknesses in AI applications. The inclusion of AI-specific attributes, such as explainability and fairness, provided a more comprehensive assessment than traditional models.

When compared with traditional models like ISO/IEC 25010, the proposed model was more effective in addressing the unique challenges of AI applications, such as transparency and adaptability [9][15].

Challenges in implementing the proposed model include the subjective nature of some quality attributes, such as fairness, and the difficulty in defining measurable metrics for continuous learning [14].

8. Conclusion

The paper presents a novel software quality model tailored specifically for AI applications, addressing both traditional and AI-specific quality attributes. The proposed model aims to guide developers and stakeholders in assessing and improving the quality of AI systems, ensuring reliability, transparency, and ethical alignment. Future work should focus on refining the quality attributes further, developing specific metrics, and applying the model to various types of AI applications.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

9. References

1. Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency (FAT), 149-159.
2. Boehm, B. W., Brown, J. R., & Lipow, M. (1976). Quantitative Evaluation of Software Quality. Proceedings of the 2nd International Conference on Software Engineering.
3. McCall, J. A., Richards, P. K., & Walters, G. F. (1977). Factors in Software Quality. General Electric Company, National Technical Information Service.
4. Doshi-Velez, F., & Kim, B. (2017). Towards a Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608.
5. Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review.
6. Garner, S. R., Zhang, H., & Brown, J. K. (2021). Explaining AI Decisions: A Survey of Explainable AI (XAI) Approaches. Journal of Artificial Intelligence Research, 68, 1-45.
7. Gezici, B., Tarhan, A.K. (2022). Systematic literature review on software quality for AI-based software. Empir Software Eng 27, 66 (2022)
8. Goodfellow, I., Shlens, J., & Szegedy, C. (2015). Explaining and Harnessing Adversarial Examples. International Conference on Learning Representations (ICLR).

9. ISO/IEC 25010:2011. Systems and Software Engineering — Systems and Software Quality Requirements and Evaluation (SQuaRE) — System and Software Quality Models. International Organization for Standardization (ISO), 2011.
10. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.
11. Wang X., Peng Y, Lu L., Lu Z., Bagheri M. & Summers R. M. (2017), ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA.
12. Vokinger, K.N., Feuerriegel, S. & Kesselheim, A.S. Mitigating bias in machine learning for medicine. Commun Med 1, 25 (2021). https://doi.org/10.1038/s43856-021-00028-w
13. Miller T (2019), Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence,Volume 267, 2019, Pages 1-38, ISSN 0004-3702,
14. Binns, R. (2018). Fairness in Machine Learning: Lessons from Political Philosophy. Proceedings of the 1st Conference on Fairness, Accountability and Transparency, in Proceedings of Machine Learning Research, 81:149-159
15. Zachary C. Lipton. 2018. The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery. Queue 16, 3 (May-June 2018), 31–57.
16. Forough Poursabzi-Sangdeh, Daniel G. Goldstein, Jake M. Hofman, Jennifer Wortman Vaughan, & Hanna Wallach. (2021). Manipulating and Measuring Model Interpretability.