



ConBRepro

XIII CONGRESSO BRASILEIRO DE ENGENHARIA DE PRODUÇÃO



IA nas Engenharias

29 nov. a 01
de dezembro 2023

Elucidando o Uso de Técnicas de Inteligência Artificial na Avaliação da Qualidade de Software Por Meio de Uma Revisão Sistemática da Literatura

Diane Lenhart

Programa de Pós-Graduação em Engenharia de Produção e Sistemas – PPGEPS,
Universidade Tecnológica Federal do Paraná – Câmpus Pato Branco

Fernando Jose Avancini Schenatto

Programa de Pós-Graduação em Engenharia de Produção e Sistemas – PPGEPS,
Universidade Tecnológica Federal do Paraná – Câmpus Pato Branco

Flavio Trojan

Programa de Pós-Graduação em Engenharia de Produção e Sistemas – PPGEPS,
Universidade Tecnológica Federal do Paraná – Câmpus Pato Branco

Resumo: Nos últimos anos, as Técnicas de Inteligência Artificial (IA) têm demonstrado sua eficácia em uma ampla gama de setores. No setor de software, a avaliação da qualidade é uma etapa importante, para a qual alguns modelos de análise baseados em IA estão sendo propostos. Hipoteticamente, entende-se que técnicas específicas de IA podem ser mais adequadas para tipos específicos de modelos de avaliação, considerando-se sua aderência, por exemplo, ao tipo de avaliação pretendida e ao tipo de software a ser avaliado. Buscando elucidar esses aspectos, o objetivo deste artigo foi analisar trabalhos onde houve o uso da IA na proposição de modelos de Avaliação da Qualidade de Software (AQS), de forma a identificar e apresentar os indicadores metodológicos e bibliométricos, além de uma discussão sobre o assunto. Para isso, foi realizada uma Revisão Sistemática da Literatura – RSL com base na declaração PRISMA. Partindo de buscas em 4 bases científicas, foram encontrados 1.541 artigos, sendo selecionados 16 para as análises. Do portfólio de trabalhos selecionados, foi identificada uma predominância de autores chineses, com 48,33% do total. Quanto aos modelos para a realização da AQS, foi constatado que 12 referem-se a novas proposições e 4 foram baseados em modelos previamente existentes. As técnicas de IA mais utilizadas foram Redes Neurais Artificiais, *Machine Learning* e Algoritmo Genético. Outras técnicas foram k-NN, Lógica *Fuzzy* e Redes *Baysianas*. Na maior parte das aplicações analisadas, a IA se mostrou eficaz na solução de problemas dentro da AQS.

Palavras-chave: Avaliação da Qualidade de Software (AQS), Inteligência Artificial (IA), Revisão Sistemática da Literatura (RSL).

Elucidating the Use of Artificial Intelligence Techniques in Software Quality Assessment Through a Systematic Literature Review

Abstract: In recent years, Artificial Intelligence (AI) techniques have demonstrated their effectiveness in a wide range of sectors. In the software sector, quality assessment is an important step, for which some AI-based analysis models are being proposed. Hypothetically, it is understood

that specific AI techniques may be more suitable for specific types of evaluation models, considering their adherence, for example, to the type of evaluation intended and the type of software to be evaluated. Seeking to elucidate these aspects, the objective of this article was to analyze works where AI was used in proposing Software Quality Assessment (AQS) models, in order to identify and present methodological and bibliometric indicators, in addition to a discussion on the subject. To this end, a Systematic Literature Review – RSL was carried out based on the PRISMA statement. Based on searches in 4 scientific databases, 1,541 articles were found, 16 of which were selected for analysis. From the portfolio of selected works, a predominance of Chinese authors was identified, with 48.33% of the total. Regarding the models for carrying out AQS, it was found that 12 refer to new propositions and 4 were based on previously existing models. The most used AI techniques were Artificial Neural Networks, Machine Learning and Genetic Algorithms. Other techniques were k-NN, Fuzzy Logic and Bayesian Networks. In most of the applications analyzed, AI proved to be effective in solving problems within AQS.

Keywords: Software Quality Assessment (SQA); Artificial Intelligence (AI); Systematic Literature Review (RSL).

1. Introdução

Nos produtos de software, os fatores invisibilidade e complexidade aumentam a dificuldade do processo de avaliação da qualidade, fazendo com que a avaliação objetiva e científica seja importante e tornando a abordagem do assunto um tópico de pesquisa. Nesse sentido, a Avaliação da Qualidade de Software (AQS) trata da quantificação das características de um produto ou processo de software (HU *et al.*, 2021; YI, 2013).

A AQS pode ser realizada durante o processo de desenvolvimento dos produtos, na avaliação do produto final e na qualidade em uso. Um pré-requisito essencial para a avaliação durante o desenvolvimento de software é a capacidade de modelar e avaliar objetivamente sua qualidade ao longo do ciclo de vida do projeto (TRENDOWICZ; KOPCZYNSKA, 2014). Já na avaliação do produto final são usados modelos de qualidade os quais fornecem uma base para especificar requisitos para avaliação (BEHKAMAL; KAHANI; AKBARI, 2009). Por outro lado, as características da qualidade em uso podem ser medidas para avaliar a qualidade do software e apontar problemas em sua produção, pois refletem a qualidade dos sistemas percebida pelos usuários finais. Entre os atributos externos da qualidade de software estão a manutenibilidade e usabilidade, afetados pela experiência do usuário (PEREIRA; POMBO; OUHBU, 2022; SOMMERVILLE, 2016).

Vários modelos e estruturas têm sido propostos ao longo dos anos para a realização do processo de avaliação (SIAVVAS; CHATZIDIMITRIOU; SYMEONIDIS, 2017; KARA; LAMOUCI; CHERIF, 2016). Os modelos de estimativa de qualidade de software estabelecem uma relação entre as características não mensuráveis e os atributos mensuráveis. No entanto, esses modelos são difíceis de generalizar e reutilizar em um software novo ou inédito, pois sua precisão se deteriora significativamente (AZAR; 2010). Apesar do esforço na criação de modelos e normas, a subjetividade da noção de qualidade dificulta a concepção e o desenvolvimento de um mecanismo universalmente aceito para o processo. Devido a isso, a pesquisa contemporânea vem buscando mecanismos capazes de produzir modelos de qualidade de software que possam ser facilmente ajustados e personalizados às necessidades do usuário (SIAVVAS; CHATZIDIMITRIOU; SYMEONIDIS, 2017).

Em contrapartida, a Inteligência Artificial (IA) alcançou um sucesso substancial em diversos campos. Máquinas desenvolvidas com a IA são capazes de aprender sozinhas e até de coletar dados, por exemplo (STAHL, 2021). A IA possui como pilares os dados, algoritmos e poder computacional. Entretanto, nas aplicações da vida real, é necessário levar outros cenários em consideração (HUAWEI; 2021). Segundo Stahl (2021), uma revisão recente da literatura sobre a IA feita pela editora acadêmica Elsevier em 2018, sugere que há uma

série de conceitos-chave e campos de pesquisa que constituem a disciplina acadêmica de IA. Em uma amostra de 600.000 documentos analisados em relação a 800 palavras-chave, o relatório classificou as publicações de IA em 7 grupos, sendo eles pesquisa e otimização, sistemas difusos, planejamento e tomada de decisão, processamento de linguagem natural e representação de conhecimento, visão computacional, aprendizado de máquinas e raciocínio probabilístico e redes neurais. Segundo Huawei (2021), acredita-se que no futuro a IA será incorporada em todas as esferas, desde automóveis, finanças, bens de consumo, saúde, educação, manufatura, comunicações, energia, entre outros.

Não foram encontrados artigos científicos que fizessem uma revisão completa e atual de modelos de avaliação utilizando técnicas de IA. No entanto, Ai-Jamimi e Ahmed (2013) realizaram um estudo bibliográfico abordando a quantificação dos parâmetros que afetam a qualidade do software e o uso técnicas e *Machine Learning* para prever a qualidade de software. Os autores concluíram que um problema comum em todos os estudos pesquisados é a transparência do modelo. A hipótese considerada neste artigo é que cada técnica de IA existente possa ser mais aderente a tipos específicos de modelos de avaliação, considerando-se os objetivos de avaliação e o tipo de software a ser avaliado. Além disso, existem várias formas de elaboração desses modelos, onde podem ser utilizadas técnicas já existentes como suporte de elaboração de requisitos. Nesse sentido, enunciou-se os seguintes questionamentos:

- RQ1: Quais técnicas de IA mais estão sendo utilizadas na AQS?
- RQ2: Quais modelos para AQS foram elaborados com a IA?
- RQ3: Quais resultados vêm sendo obtidos dentro do campo de IA na AQS?
- RQ4: Quais as implicações observadas dentro do campo de IA na AQS?

Considerando o que foi abordado sobre a AQS, a versatilidade de aplicações da IA, a lacuna identificada a respeito da intersecção desses temas e afim de buscar responder as perguntas de pesquisa, estabeleceu-se como objetivo analisar a ocorrência de aplicações de IA à AQS, discutindo os modelos e técnicas utilizadas bem como os resultados alcançados e suas implicações por meio de uma Revisão Sistemática da Literatura – RSL.

Este artigo está dividido em 4 seções. A primeira contextualiza o tema de pesquisa, bem como os objetivos e a justificativa. A segunda é composta pela metodologia, na qual foram apresentadas as ferramentas utilizadas para execução do trabalho como um todo e o processo de refinamento da busca. Na terceira seção são apresentados os resultados encontrados, contendo as respostas relativas às questões de pesquisa feitas e discussões. Na quarta seção são apresentadas as conclusões.

2. Metodologia

Para que o trabalho fosse feito de forma sistemática em sua completude, optou-se pelo uso do “*Prisma Statement*”, em português “Declaração Prisma” da PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), que consiste em uma lista de verificação com instruções a serem seguidas na realização de revisões bem como um fluxograma a ser preenchido durante as etapas de refinamento de busca (PRISMA, 2023). A versão utilizada foi a de 2020.

As buscas por artigos científicos foram realizadas em bases eletrônicas disponíveis pelo portal da CAPES. As bases utilizadas para a pesquisa foram: ACM Digital library, IEEEXplore, ScienceDirect (Elsevier) e SCOPUS (Elsevier).

Os eixos definidos foram Inteligência Artificial (IA) e Avaliação da Qualidade de Software (AQS). Para o primeiro eixo, definiram-se as palavras-chave “*artificial intelligence*”, “*machine learning*”, “*deep learning*” e “*neural networks*”. Verificou-se que a AQS é tratada

de diversas formas na literatura, sendo assim, definiram-se palavras-chave com combinações estratégicas, conforme mostra o Quadro 1. Com o auxílio de operadores booleanos, a expressão de busca definida foi a seguinte: (((“software quality assessment”) OR (“quality” AND “software assessment”) OR (“evaluation model” AND “software quality”)) AND (“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural networks”)).

As buscas nas bases científicas foram limitadas para artigos de periódicos, jornais e congressos, além dos anos de publicação, sendo de 2000 a 2023.

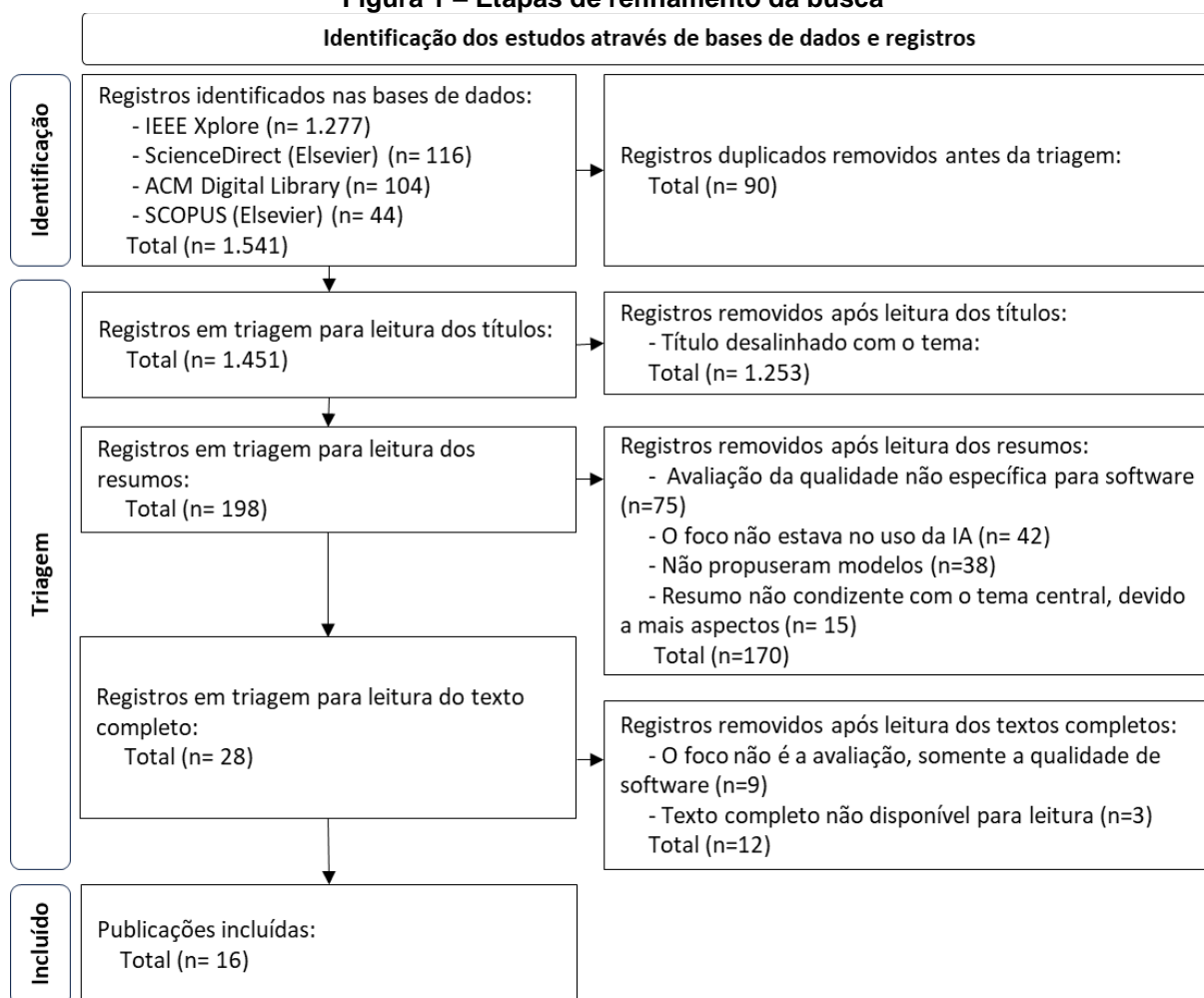
Quadro 1 – Combinações de busca nas bases

AQS				IA	
“software quality assessment”	-	-	AND	“artificial intelligence” OR “machine learning” OR “deep learning” OR “neural networks”	
“quality”	AND	“software assessment”	AND		
“evaluation model”	AND	“quality”	AND		

Fonte: Elaborado pelo autor (2023)

As 4 bases selecionadas para busca retornaram 1.541 artigos no total, sendo 1.277 da IEEEXplore, 116 da ScienceDirect (Elsevier), 104 da ACD Digital Library e 44 da SCOPUS (Elsevier). A etapa de refinamento da pesquisa está esquematizada na Figura 1.

Figura 1 – Etapas de refinamento da busca



Fonte: Elaborado pelo autor (2023)

Na etapa de identificação, as referências dos artigos foram exportadas para o software *Mendeley*®, sendo excluídos 90 documentos duplicados, restando 1.451 artigos. A execução das próximas etapas se deu de forma subjetiva pelos autores deste artigo. Na

etapa de triagem, determinou-se que 198 dos 1.451 títulos eram condizentes com o tema de pesquisa, sendo eliminados 1.253 artigos. Essa é uma exclusão significativa e se deve ao fato de que muitos dos artigos eram voltados para áreas específicas, como avaliação da qualidade de imagem, qualidade do serviço, qualidade da água e bioquímica e genética. Com a leitura dos resumos dos 198 artigos restantes, identificou-se que em 75 dos artigos a avaliação não foi feita especificamente para o setor de software, 42 não usaram alguma técnica de IA, 38 não propuseram novos modelos e 15 não estavam alinhados com o tema devido a mais aspectos, sendo estes eliminados, restando 28 artigos. Ainda na etapa de triagem, foram lidos os 28 textos completos selecionados, no qual 9 foram eliminados. Nesse caso, os artigos continham os assuntos qualidade de software e IA, mas o foco não era a avaliação. Os 3 trabalhos restantes foram eliminados devido o texto não estar disponível para leitura. Dessa forma, o total de artigos incluídos para análise foi 16.

3. Resultados e discussões

3.1 Portfólio selecionado, informações gerais e técnicas utilizadas

No Quadro 2 estão dispostos os títulos dos 16 artigos selecionados para o portfólio. Nas colunas, “Ref.” representa o número de referência de artigo, que será utilizada nas próximas seções, “Ve.” Representa o veículo de publicação no qual os artigos publicados em periódicos estão representados pela letra “P” e os publicados em conferências pela letra “C” e a coluna “Cit” refere-se à quantidade de citações levantadas no Google Acadêmico. Outras informações são o ano de publicação e o(s) autor(es).

Quadro 2 – Artigos selecionados para o portfólio

Ref.	Ano	Ve.	Título	Autor(es)	Cit.
1	2023	C	<i>Software Risk Ranking Assessment Model using Fuzzy Decision-Making Trial and Evaluation Laboratory</i>	Suresh, Kavitha e Poonkodi	0
2	2021		<i>Software Defect Prediction Method Based on Transformer Model</i>	Zheng, Than e Liu	3
3			<i>Qunomon: A FAIR testbed of quality evaluation for machine learning models</i>	Narita <i>et al.</i>	1
4			<i>Marine TT&C Software Quality Evaluation based on AHP and FCE</i>	Su <i>et al.</i>	0
5	2020	P	<i>A case study for software quality evaluation based on SCT model with BP neural network</i>	Yan <i>et al.</i>	3
6	2019	C	<i>Evaluate Software Quality by Learning from Historical Data</i>	Zong <i>et al.</i>	0
7	2017		<i>Comprehensive Evaluation of Software Quality Based on LM-BP Neural Network</i>	Wang <i>et al.</i>	2
8			<i>A Software Development Project Quality Analysis Model Based on HMM-FNN</i>	Yanxu <i>et al.</i>	1
9	2016		<i>Simulating reliability of the robotic system software on the basis of artificial intelligence</i>	Sheptunov <i>et al.</i>	19
10	2015		<i>Machine Learning Approach for Quality Assessment and Prediction in Large Software Organizations</i>	Rana e Staron	21
11	2014		<i>Qualitygate SourceAudit: A tool for assessing the technical quality of software</i>	Bakota <i>et al.</i>	30
12	2013		<i>Quality assessment of software as a service on cloud using fuzzy logic</i>	Baliyan e Kumar	39
13			<i>A new software maintainability evaluation model based on multiple classifiers combination</i>	Ye, Zhu e Wang	20
14	2010	P	<i>A genetic algorithm for improving accuracy of software quality predictive models: A search-based software engineering approach</i>	Azar	11
15	2008	C	<i>2-D software quality model and case study in software flexibility research</i>	Zang, Li e Gao	15
16	2002		<i>Software Quality Analysis with the use of Computational Intelligence</i>	Reformat, Pedrycz e Pizzi	67

Fonte: Elaborado pelo autor (2023)

Em relação aos anos das publicações, há destaque para o ano de 2021, 2017 e 2013. Entre 2000 e 2012 o total de artigos publicados corresponde a 18,75% e entre 2013 e 2023, o total é 81,25%. Constatou-se que 87,5% dos artigos foram apresentados em conferências. Destacou-se o artigo intitulado “*Software Quality Analysis with the use of Computational Intelligence*” dos autores Reformat, Pedrycz e Pizzi, com 67 citações, sendo o artigo mais citado. É importante observar que o artigo é do ano de 2002, enquanto muitos dos artigos são mais atuais, o que pode interferir na quantidade de citações.

A China sobressaiu-se como o país com a maior quantidade de autores, com o percentual de 48,33%, seguido de Japão (15%) e Rússia (10%). Segundo Shen (2018), os dados oficiais do Ministério da Indústria e Tecnologia da Informação da China mostram que o rendimento dos engenheiros envolvidos na indústria de software vem aumentando significativamente, possuindo perspectivas positivas para o futuro.

A classificação dos modelos (novos ou de melhoria), o tipo de software em questão (específico ou não), a técnica IA utilizada, a abordagem AQS, a classificação quanto o modelo ser testado ou não é a referência do artigo (conforme indicado no Quadro 2), são apresentados no Quadro 3.

Uma análise importante neste trabalho para cumprimento dos objetivos definidos à priori foi a análise das técnicas de IA utilizadas, que responde a questão de pesquisa “RQ1: Quais técnicas de IA mais estão sendo utilizadas na AQS?”, definida na introdução. Em análise, foram identificadas como 20 o total de técnicas utilizadas considerando os 16 trabalhos do portfólio.

A técnica de maior frequência de uso foram as Redes Neurais de variados tipos, sendo utilizada em 5 trabalhos. Essas técnicas incluem a Rede Neural Artificial sem denominação específica e redes específicas, como a Rede Neural BP, Rede Neural LB-BP e Rede Neural Convolutiva. A segunda técnica de maior frequência foi *Machine Learning*, utilizada em 4 trabalhos.

A Lógica Fuzzy apareceu em 3, sendo elas a Lógica Fuzzy sem denominação específica, a *Fuzzy Comprehensive Evaluation* e *Fuzzy DEMATEL*.

Algoritmos Genéticos foram utilizados em 3 trabalhos e o algoritmo k-NN em 2. Outras técnicas foram Redes Neurais Fuzzy, *Deep Learning* e Redes Bayesianas.

Respondendo à questão de pesquisa “RQ2: Quais modelos para AQS foram elaborados com a IA?”, foram analisados os tipos de modelos elaborados. É importante salientar que todos os trabalhos estão dentro do campo da AQS, entretanto, alguns são mais específicos a alguns objetivos como para o desenvolvimento, previsão, teste, precisão da predição, confiabilidade, previsão de defeitos, manutenibilidade e riscos no projeto de software.

Em 12 dos 16 trabalhos do portfólio, os autores elaboraram novos modelos para avaliação da qualidade de software e, em 4 deles, o propósito foi melhorar modelos já existentes. Dos 12 trabalhos onde foram elaborados novos modelos, 10 foram feitos para serem aplicados a qualquer tipo de software e em 2 deles o modelo foi feito para softwares específicos, sendo um para aviação e outro para o software SaaS. Quanto aos modelos de melhoria, 3 foram elaborados sem ser para algum software específico e um deles para o software NASA 93-COCOMO.

Em 4 das obras, observou-se o uso de modelos específicos de qualidade. Nesses casos, os modelos foram utilizados como embasamento para o entendimento de conceitos de qualidade, não sendo esses utilizados para a melhoria desses modelos. Os modelos mais utilizados referem-se às normas ISO, utilizada em três trabalhos, sendo esses as normas ISO 9126, ISO 25010 e ISO 15939.

Em 93,75% dos trabalhos os autores realizaram o teste do modelo, ou seja, o modelo foi aplicado para verificação da eficácia.

Quadro 3 – Detalhamento das técnicas IA e abordagens AQS do portfólio

Modelo	Tipo de software	Abordagem AQS	Técnica IA	Modelo AQS	Teste	Ref
Novo	Não Específico	Avaliação da qualidade de software	Rede Neural BP	SCT	S	5
			Rede Neural LM-BP	ISO/IEC 9126		7
			Rede Neural BP, <i>Machine Learning</i> e KNN	-		11
			Classificadores Genéticos			16
		Avaliação da qualidade do desenvolvimento de software	Redes Neurais <i>Fuzzy</i>	8		
		Avaliação e previsão da qualidade de software	<i>Machine Learning</i>	ISO 15939		10
		Avaliação e teste de qualidade		3		
		Precisão da predição da qualidade	Algoritmo genético	-		14
	Manutenibilidade de software	13				
	Aviação	Avaliação da qualidade de software	<i>Machine Learning</i> , Algoritmo Genético e k-NN	ISO/IEC 25010	N	6
	SaaS		Lógica <i>Fuzzy</i>	12		
	TT&C Marítimo		<i>Fuzzy Comprehensive Evaluation</i>	4		
	Melhoria	Não Específico	Avaliação da confiabilidade de software	Redes Neurais Artificiais	-	S
Previsão de defeitos de software			Rede Neural Convolutiva e <i>Deep Learning</i>	2		
Avaliação e previsão da qualidade de software			Redes <i>Baysianas</i>	15		
NASA 93-COCOMO		Avaliação de classificação de risco de software	<i>Fuzzy DEMATEL</i>	1		

Fonte: Elaborado pelo autor (2023)

3.2 Uso de técnicas de inteligência artificial na elaboração de modelos de avaliação da qualidade de software

Na proposição de modelos de avaliação da qualidade de software, um processo importante é a definição de um embasamento teórico sobre o assunto dentro do campo da qualidade. Nos artigos do portfólio selecionado, muitos autores realizaram isso fazendo uma revisão teórica sobre o assunto em interesse, mencionando autores e teorias diversas. Esse processo pode ser realizado com o uso de modelos prontos, como é o caso do uso de normas ISO pelos autores. No artigo de referência 10, por exemplo, os autores utilizaram o modelo de informação de medição ISO 15939, mostrando como diferentes métricas de software podem ser usadas na construção de um modelo para avaliação e previsão da qualidade de software em grandes organizações. O uso do modelo SCT (*Scope, Cust and Time*) pelos autores Yan *et al.* (2020), representa uma possível versatilidade existente no uso desses modelos. Nesse caso, considera-se que de acordo com as diferentes características dos projetos de software, como volume de código, estrutura técnica e cultura da equipe, os índices de avaliação quantitativa também têm pesos diferentes sobre o efeito

do padrão de qualidade utilizado (escopo, custo e tempo), em todo o sistema de avaliação de proposta.

Como mencionado na seção 3.1, Redes Neurais Artificiais foi a técnica de IA mais utilizada, dentre as suas variações. No artigo de referência 9, os autores elaboraram um tipo especial de rede neural, uma rede neural de camadas verticalmente. O modelo proposto foi desenvolvido para a estimativa da confiabilidade de software em todos os estágios de ciclo de vida e para determinar o valor previsto da taxa de falhas com base nos dados de confiabilidade. Também foi elaborado um modelo matemático para extrair a relação de influência direta e indireta para esses atributos. No artigo de item 7, os autores construíram um método abrangente de qualidade de software baseado na rede neural LM-BP, onde os resultados experimentais mostraram que o algoritmo pode melhorar a velocidade de treinamento da rede e a precisão é alta. A Rede Neural Convolucional foi utilizada na proposição de um método de previsão de defeitos de software, onde os autores basearam-se em modelo existente denominado “*Transformer*”.

No campo de teste de software relacionados à *Machine Learning*, Narita *et al.* relataram que, embora vários métodos de teste de *Machine Learning* tenham sido propostos, eles não foram totalmente adaptados aos processos de desenvolvimentos de componentes de *Machine Learning* devido às suas configurações de teste individuais. Devido a isso, os autores elaboraram um sistema de teste denominado Qunomon, que implanta e executa métodos de avaliação de qualidade *Machine Learning* em um sistema local. O uso de *Machine Learning* pode ser combinado com outras técnicas de IA, como k-NN e Algoritmo Genético, que é o caso do artigo de referência 6, onde o algoritmo de *Machine Learning* ofereceu suporte ao treinamento incremental. O k-NN foi usado para estabelecer a relação entre a confiabilidade e medição de código. Para Zong *et al.* (2019), uma situação negativa pode ocorrer nos casos de avaliações de dados históricos utilizando *Machine Learning* quando a avaliação da qualidade é realizada com dados históricos de um mesmo software. Segundo os autores, nesses casos não se pode avaliar com precisão o software de outras empresas e outros campos, porém, os erros subjetivos podem ser reduzidos pelos fatores humanos dos avaliadores, aumentando medidas externas de software e aumento da quantidade de dados de treinamento.

Segundo Rasyid, Mubtadai e Abdulrokhim (2019), Algoritmos Genéticos são elaborados para encontrar estimativas de conclusão para problemas de otimização e busca. No artigo de referência 14, essa técnica foi utilizada para otimizar modelos preditivos utilizados para estimar características de qualidade de software, como acoplamento, tamanho e complexidade. O Algoritmo Genético também se mostrou eficaz na elaboração de um modelo com foco em manutenibilidade de software. Neste artigo (referência 13), foi realizada uma análise a respeito dos atuais métodos de avaliação de manutenibilidade. Para eles, muitos modelos apresentam erros específicos nas etapas de pré-processamento de dados, treinamento e avaliação do modelo. A partir desses erros observados, pensou-se em melhorias a serem aplicadas na elaboração do novo modelo, dividindo-o em três partes, sendo eles a seleção de atributos, treinamento e interpretação do modelo. O modelo, denominado SMEM-MCC, mostrou capacidade em prever a manutenibilidade, provada através de uma análise experimental.

No artigo de referência 6, o Algoritmo Genético foi utilizado para avaliar a confiabilidade de software. Isso mostra a versatilidade da técnica dentro da AQS, onde pode ser mencionado também a sua eficácia nesses contextos, sendo que em todos os casos os resultados foram positivos, segundo os autores. Segundo os autores do artigo Reformat, Pedrycz e Pizzi (2002), com o uso dos classificadores genéticos um gerente de sistema pode prever a qualidade dos objetos de software e identificar os objetos que não são de qualidade para reavaliação e possível previsão”. Nesse caso os classificadores foram utilizados para realizar a classificação de objetos de software com relação à sua qualidade.

Técnicas de IA podem ser usadas para a elaboração de modelos com auxílio de juntamente com técnicas de outros campos, como é o caso do artigo de Su *et al.* (2020). Os autores construíram um índice de avaliação baseado no AHP (*Analytic Hierarchy Process*), juntamente com a utilização do *Fuzzy Comprehensive Evaluation*, especificamente para a avaliação da capacidade de execução de tarefas do software TT&C marítimo. Em aplicação real, os autores mostraram que a utilização das técnicas foi viável.

Modelos *Fuzzy* foram usados em 2 softwares específicos, sendo o SaaS e COCOMO. Não se encontrou evidência de um significado para isso em termos do uso das técnicas. Segundo Baliyan e Kumar (2013), o Software como Serviço (SaaS) é um modelo de entrega para computação em nuvem, onde a prioridade do software pelo provedor de SaaS é isolada de seu uso pelo cliente de SaaS. No modelo elaborado para o software COCOMO, utilizou-se o *Fuzzy DEMATEL* melhorado, onde os autores concluíram que sua eficiência melhorou significativamente através das modificações feitas. Na junção das técnicas de Redes Neurais e Lógica *Fuzzy*, denominado Redes Neurais *Fuzzy*, Yanxu *et al.* (2020) elaboraram um modelo capaz de avaliar a qualidade do projeto de desenvolvimento de software com precisão. O modelo pode fornecer uma referência eficaz para otimizar a estrutura do projeto de desenvolvimento.

A utilização de métodos *baysianos* também foi identificada. A proposta observada no portfólio em relação a essa técnica foi um modelo de qualidade denominado software 2-D, que sintetiza o modelo de qualidade decomposto tradicional e o modelo de qualidade de relação para recursos, sub-recursos e métricas de qualidade de software presentes e abrangentes. Nesse caso, a teoria das Redes *Baysianas* é utilizada para construir relações de incerteza.

4. Conclusões

Neste artigo foi realizada uma Revisão Sistemática da Literatura – RSL buscando explanar a utilização de técnicas de Inteligência Artificial para a realização da Avaliação da Qualidade de Software. O intuito foi identificar modelos elaborados através dessas técnicas, analisando a eficácia da utilização em cada caso além de fornecer um panorama geral sobre o tema.

De um total de 1.541 artigos, foram selecionados 16 para análise, sendo que a maior quantidade de publicações está entre os anos de 2013 e 2023 e 87,5% foram apresentados em conferências. A China sobressaiu-se como o país com a maior quantidade de autores, com o considerável percentual de 48,33% do total. O artigo com o maior número de citações é denominado “*Software Quality Analysis with the use of Computational Intelligence*” dos autores Reformat, Pedrycz e Pizzi. Constatou-se que em 12 dos trabalhos os autores objetivaram a elaboração de novo modelos para a realização da AQS em softwares, e em 4, o foco foi a melhoria de modelos já existentes já existentes. A técnica de IA mais utilizada foram as Redes Neurais Artificiais, seguido de *Machine Learning* e Algoritmos Genéticos. Outras técnicas são voltadas à Lógica *Fuzzy*, k-NN e Redes *Baysianas*.

Na maior parte dos modelos elaborados a aplicação de IA se mostrou eficaz na obtenção dos resultados de pesquisa. Aspectos negativos foram citados quanto a utilização de dados históricos para a construção de método para AQS, onde salientou-se que a elaboração de modelos de uma mesma empresa dificulta a utilização em outros tipos de software. Algo que se percebeu durante a realização na etapa de descrição de conteúdo, é que existiram diversas formas e caminhos para a apresentação das ideias, devido à complexidade dos dois temas de modo geral e dos artigos selecionados no portfólio. Nesse aspecto, é inevitável que a conjunção do conteúdo esteja vinculada à subjetividade de alguma forma. Todavia, buscou-se dar ênfase no que responderia mais diretamente os objetivos elaborados.

De um modo geral, as questões de pesquisa puderam ser respondidas através das análises bibliométricas, metodológicas e de conteúdo. Expecta-se que este trabalho beneficie aprendizes e entusiastas no assunto. Também se entende que as informações relatadas, principalmente quanto ao uso específico e a eficácia dos métodos, poderão contribuir significativamente para que engenheiros de software, desenvolvedores e profissionais da qualidade se beneficiem, partindo para uma melhor visão de utilização das técnicas de IA na elaboração e construção de modelos de AQS.

Como trabalhos futuros recomenda-se a aplicação dos métodos de IA apresentado pelos autores do portfólio que relataram resultados positivos. Partindo das informações apresentadas pode-se ampliar e elaborar novos modelos, por meio do cruzamento de informações. Outra recomendação é a realização de uma nova RSL. Nesse caso, portfólio obtido pode ser modificado com um processo mais robusto na escolha das palavras-chave. Isso pode ser obtido encontrando alguma forma de especificar subáreas da qualidade para a AQS e, quanto a IA, mencionado suas técnicas de forma mais específica. Como dito, os dois temas são amplos, sendo assim, alterações na busca podem modificar, ampliar e até melhorar o portfólio de artigos encontrados e conseqüentemente a análise de conteúdo como um todo.

Referências

AI-JAMIMI, H. A.; AHMED, M. Machine learning-based software quality prediction models: State of the art. In: INTERNATIONAL CONFERENCE ON INFORMATION SCIENCE AND APPLICATIONS, 1., 2013, Pattaya. **Anais eletrônicos...** Pattaya, Thailand, 2013. p. 1-1.

AZAR, D. A genetic algorithm for improving accuracy of software quality predictive models: A search-based software engineering approach. **International Journal of Computational Intelligence and Applications**, v.9, n. 2, p. 125-136, June 2010.

BALIYAN, N.; KUMAR, S. Quality assessment of software as a service on cloud using fuzzy logic. In: 2nd IEEE INTERNATIONAL CONFERENCE ON CLOUD COMPUTING IN EMERGING MARKETS, 2013. Índia. **Anais eletrônicos...** India, 2013.

BAKOTA, T.; HEGEDUS, P.; SIKET, I.; LADÁNYI, G.; FERENC, R. Qualitygate SourceAudit: A tool for assessing the technical quality of software. In: SOFTWARE EVOLUTION WEEK - IEEE CONFERENCE ON SOFTWARE MAINTENANCE, REENGINEERING, AND REVERSE ENGINEERING, 2014, Antwerp. **Anais eletrônicos...** Antwerp, Belgium, 2014. p. 440-445.

BEHKAMAL, B.; KAHANI, M.; AKBARI, M. K. Customizing ISO 9126 quality model for evaluation of B2B applications. **Information and Software Technology**, v. 51, p. 599-609, 2009.

HU, X.; LI, H.; ZHANG, M.; QU, H. A Novel Software Quality Evaluation Method Based on Fuzzy Comprehensive Evaluation Model. In: IEEE 5TH ADVANCED INFORMATION TECHNOLOGY, ELECTRONIC AND AUTOMATION CONTROL CONFERENCE, 5., 2021, China. **Anais eletrônicos...** China, 2021. p. 2275-2279.

HUAWEI, T. C. Ltd. **Artificial Intelligence Technology**. Hangzhou, China: Springer, Posts & Telecom Press, 2021.

KARA, M.; LAMOUCHE, O.; CHERIF, A. R. Ontology Software Quality Model for Fuzzy Logic Evaluation Approach. **Procedia Computer Science**, v. 86, p. 637-641, 2016.

PEREIRA, L. S.; POMBO, N.; OUHBI, S. Software quality: Application of a process model for quality-in-use assessment. **Journal of King Saud University – Computer and Information Sciences**, v. 34, p. 4626-4634, 2022.

PRISMA – **TRANSPARENT REPORTING of SYSTEMATIC REVIEWS and META-ANALYSES**. 2023. Disponível em: < <http://www.prisma-statement.org/>> Acesso em: 20 jun. 2023.

RANA, R.; STARON, M. Machine Learning Approach for Quality Assessment and Prediction in Large Software Organizations. In: PROCEEDINGS OF THE IEEE INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND SERVICE SCIENCES, 6., 2015. **Anais eletrônicos...** 2015, n. 7339243. p.1098-110125.

RASYID, M. U. H. A.; MUBTADAI, N. R.; ABDULROKHIM, J. Performance Analysis LEACH Based Genetic Algorithm In Wireless Sensor Network. In: INTERNATIONAL SEMINAR ON APPLICATION FOR TECHNOLOGY OF INFORMATION AND COMMUNICATION, 1., 2019, Semarang. **Anais eletrônicos...** Semarang, Indonesia, 2019. p. 394-399.

REFORMAT, M; PEDRYCZ, W.; PIZZI, N. J. Software quality analysis with the use of computational intelligence. In: IEEE INTERNATIONAL CONFERENCE ON FUZZY SYSTEMS, 1., 2002, Honolulu. **Anais eletrônicos...** Honolulu, USA, 2002, v. 2., p. 1156-1161.

NARITA, K.; AKITA, M.; KIM, K-S; IWASE, Y.; WATANAKA, Y.; NAKAGAWA, T.; ZHONG, Q. Qunomon: A FAIR testbed of quality evaluation for machine learning models. In: ASIA-PACIFIC SOFTWARE ENGINEERING CONFERENCE WORKSHOPS, 28., Taipei, 2021. **Anais eletrônicos...** Taipei, Taiwan, 2021. p. 21-24.

SHEPTUNOV, S. A.; LARIONOV, M. V.; SUHANOVA, N. V.; SALAKKHOV, M. R.; SOLOMENTSEV, Y. M.; KABAK, I. S. Simulating reliability of the robotic system software on the basis of artificial intelligence. In: IEEE CONFERENCE ON QUALITY MANAGEMENT, TRANSPORT AND INFORMATION SECURITY, INFORMATION TECHNOLOGIES, 1., 2016, Nalchik. **Anais eletrônicos...** Nalchik, Russia, 2016. p. 193-197.

SIAVVAS, M. G.; CHATZIDIMITRIOU, K. C.; SYMEONIDIS, A. L. QATCH - An adaptive framework for software product quality assessment. **Expert Systems With Applications**, v. 86, p. 350-366, 2017.

SOMMERVILLE, I. **Software Engineering**. New York, USA: Ed. Addison Wesley, 2016.

STAHL, B. C. **Artificial Intelligence for a Better Future. An Ecosystem Perspective on The Ethics of AI and Emerging Digital Technologies**. Reino Unido: Ed. SpringerBriefs In. Research and Innovation Governance, 2021.

SU, C.; LI, Y.; GUO, L.; MAO, W. Marine TT&C Software Quality Evaluation based on AHP and FCE. In: Joint International Information Technology and Artificial Intelligence Conference, 9., 2020, Chongqing. **Anais eletrônicos...** Chongqing, China, 2020, p. 72-76.

SURESH, D. K.; KAVITHA, M. V.; POONKODI, M. S. Software Risk Ranking Assessment Model using Fuzzy Decision-Making Trial and Evaluation Laboratory. In: INTERNATIONAL CONFERENCE ON SMART SYSTEMS AND INVENTIVE TECHNOLOGY, 5., 2023, Tirunelveli. **Anais eletrônicos...** Tirunelveli, India, 2023. p. 1485-1490.

TRENDOWICZ, A.; KOPCZYNSKA, S. Adapting Multi-Criteria Decision Analysis for Assessing the Quality of Software Products. Current Approaches and Future Perspectives. **Advances in Computers**, v. 93, p.153-226, 2014.

WANG, L.; GUO, L.; CHEN, L.; WANG, J.; SONG, Y.A. Comprehensive Evaluation of Software Quality Based on LM-BP Neural Network. In: INTERNATIONAL CONFERENCE ON DEPENDABLE SYSTEMS AND THEIR APPLICATIONS, 2017, Beijing. **Anais eletrônicos...** Beijing, China, 2017, p. 162-167.

YAN, B. YAO, H.; NAKAMURA, LI, -F LI.; WANG, D. A Case Study for Software Quality Evaluation Based on SCT Model With BP Neural Network. **IEEE Access**, v. 8, p. 56403-56414, 2020.

YE, F.; ZHU, X.; WANG, Y. A new software maintainability evaluation model based on multiple classifiers combination. In: INTERNATIONAL CONFERENCE ON QUALITY, RELIABILITY, RISK, MAINTENANCE, AND SAFETY ENGINEERING, 1., 2013, Chengdu. **Anais eletrônicos...** Chengdu, China, 2013. p.1588-1591.

YI, C. Research and Realization of Software Quality Fuzzy Comprehensive Evaluation Based on AHP. **Computer Measurement & Control**, v. 21, n.12, p. 3306-3308, 2013.

YUANXU, L.; CHAOWEN, H.; GUO, W. A software development project quality analysis model based on HMM-FNN. In: IEEE INTERNATIONAL CONFERENCE ON SOFTWARE ENGINEERING AND SERVICE SCIENCE, 8., 2017, Beijing. **Anais eletrônicos...** Beijing, China, 2017. p. 220-224.

ZHANG, L.; LI, L.; GAO. 2-D Software Quality Model and Case Study in Software Flexibility Research. In: INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE FOR MODELLING CONTROL & AUTOMATION, 1., 2008, Vienna. **Anais eletrônicos...** Vienna, Austria, 2008. p. 1147-1152.

ZHENG, W.; TAN, L.; LIU, C. Software Defect Prediction Method Based on Transformer Model. In: IEEE INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND COMPUTER APPLICATIONS, 1., 2021, Dalian. **Anais eletrônicos...** Dalian, China, 2021, p. 670-674.

ZONG, P.; WANG, Y; SONG, Z. Evaluate Software Quality by Learning from Historical Data. In: INTERNATIONAL CONFERENCE ON INTELLIGENT COMPUTING, AUTOMATION AND SYSTEMS, 1., 2019, Chongqing. **Anais eletrônicos...** Chongqing, China, 2019, p. 16-20.