TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)
31 Oct - 3 Nov 2023. Chiang Mai, Thailand

Artificial Intelligence in Software Testing: A Systematic Review

Mahmudul Islam
Department of Computer Science and Engineering
Independent University, Bangladesh
mahmud@iub.edu.bd

Farhan Khan
Department of Computer Science and Engineering
Independent University, Bangladesh
farhankhan_262@yahoo.com

Sabrina Alam
Department of Computer Science and Engineering
Independent University, Bangladesh
sabrina.alam@iub.edu.bd

Mahady Hasan
Department of Computer Science and Engineering
Independent University, Bangladesh
mahady@iub.edu.bd

***Abstract*—Software testing is a crucial component of software development. With the increasing complexity of software systems, traditional manual testing methods are becoming less feasible. Artificial Intelligence (AI) has emerged as a promising approach to software testing in recent years. This review paper aims to provide an in-depth understanding of the current state of software testing using AI. The review will examine the various approaches, techniques, and tools used in this area and assess their effectiveness. The selected articles for this study have been extracted from different research databases using the advanced search string strategy. Initially, 40 articles have been extracted from different research libraries. After gradual filtering finally, 20 articles have been selected for the study. After studying all the selected papers, we find that various testing tasks can be automated successfully using AI (Machine Learning and Deep Learning) such as Test Case Generation, Defect Prediction, Test Case Prioritization Metamorphic Testing, Android Testing, Test Case Validation, and White Box Testing. This study also finds that the integration of AI in software testing is making software testing activities easier along with better performance. This literature review paper provides a thorough analysis of the impact AI can have on the software testing process.**

***Index Terms*—Software Testing, Artificial Intelligence, Test Automation, Systematic Literature Review**

I. Introduction

Software testing has a crucial role in software engineering as it is essential for ensuring the quality, performance, security, and reliability of software systems. By conducting testing, developers can identify and rectify any bugs, or defects in the software, improving its overall functionality and making sure that the software satisfies customer needs and expectations. AI is a vast area , so in this paper we mainly investigate the subarea of AI which are Machine Learning (ML) and Deep Learning (DL) techniques in software testing. The field of software testing currently faces a number of challenges. As software systems grow increasingly complex, it becomes more challenging to manually test all possible scenarios. Also, traditional test automation approaches are time-consuming and complex to implement. Apart from that, keeping pace with agile development is also a challenge as it requires rapid testing. AI has the potential to address these challenges by offering optimized and effective testing strategies. The aim of this study is to gain a thorough understanding of the current state of the field of software testing automation through the use of AI. This review will examine the various methods, techniques, and tools utilized in this domain and evaluate their efficiency. The motivation for this systematic literature review stems from the potential benefits that AI can offer in the field of software testing. AI has the potential to automate the testing process and optimize testing strategies, making software testing more efficient, effective, and accessible. Moreover, AI can address the shortage of skilled testers and help keep pace with the rapid development cycles of agile development methodologies. There are several challenges in software testing that can be solved using AI. Some of these issues include manually generating test cases, test optimization, test results analysis, etc.

The following research questions have been investigated in this research study.

RQ1: Does manual testing have drawbacks?

RQ2: Can integration of AI (ML or DL techniques) in software testing help to overcome the drawbacks of manual testing?

RQ3: What software testing tasks can be automated by AI (ML or DL)?

RQ4: What techniques do researchers use to assess AI (ML or DL) when used in software testing?

In this research study, 40 articles have been screened from different research libraries but through a gradual filtering process, only 20 articles were found suitable for the study. We have structured the paper in the following way. Related works have been discussed in section 2 while the background of software testing and AI have been presented in section 3. Systematic review and the results have been presented in section 4 and 5 consecutively. In the end, conclusion is presented in section 6.

II. Related Works

They [1] proposed a deep learning model to rank test cases. In this work, they consider historical records of test case executions and based on that deep learning model rank test cases. They [2] conducted an empirical study on continuous integration testing. They found the strategy of reward function of Reinforcement learning improves the existing test case prioritization practices. They [3] developed a deep reinforcement learning technique for performing black

TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON) | 979-8-3503-0219-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/TENCON58879.2023.10322349

box testing on android apps. Their developed technique outperforms existing techniques in terms of fault identification. They [4] proposed a deep learning-based approach for prioritizing test cases from the interaction of humans with software applications. They showed that test case prioritization can perform successfully from human interactions using their proposed model. They [5] presented an approach to generate input for the graphical user interface of software applications by only capturing screenshots of applications.

They [6] proposed a machine learning-based approach to predict metamorphic relations of scientific software using graph kernels. They concluded that features extracted from graphs help to achieve a good result. They [7] presented an approach to automate test oracle mechanism using machine learning. Their proposed approach captures historical usage data and based on that generates an oracle. They [8] detected metamorphic relations using graph kernels and support vector machines (SVM). They [9] analyzed software defect prediction using machine learning algorithms. They found that linear classifier performs well compared to other algorithms. They [10] proposed an improved CNN model to predict software defects and their proposed model outperformed existing models.

III. Software Testing & Artificial Intelligence

Software Testing is a process to evaluate the software and identify defects [11]. It is crucial for software to work or perform as per requirements but it is natural having bugs or defects in software. The bugs can be generated during development, bug fixing, feature addition, code refactoring, and even during software maintenance [12]. Therefore, it is obvious for the development team to test the software under different scenarios before releasing it to the client. There are different strategies and techniques for software testing. Based on the nature of the software it would be decided which software testing technique should be used [13]. Software testing techniques are very tedious and automation comes here to ease the process. How AI can automate software testing and why it is getting more acceptance than any other technique will be discussed in this section. AI is a broad area that encompasses various subareas, and ML is one of the most prominent and widely applied subareas within AI. In this paper, we discuss software testing using Machine Learning (ML). We also focus on software testing using Deep Learning (DL).

A. Software Testing Using Machine Learning

Machine Learning (ML) is a process where machines learn from data using algorithms and can further predict or make decisions based on the data [14]. The data-centric learning approach has made Machine Learning powerful and widely accepted in different areas including the software industry. Figure 1 shows the general approach to apply Machine Learning algorithms in software testing. There are different activities in software testing like bug detection, generating test data, test case generation, test optimization, API testing, etc [**?**].

Bug Prediction using ML: Bug prediction can be performed using machine learning. ML algorithms analyze software code and predict the likelihood of future bugs in the code. For performing bug prediction, ML models need to be trained on historical data from past software projects to identify patterns. Once the model is trained, then it can predict the likelihood of bugs occurring in new code [15]. They [16] used supervised ML algorithms to predict software faults based on historical data.

Test Case Generation using ML: In software development, Test case generation from the requirement specifications document is one of the biggest challenges in software testing. Software test cases can be generated using ML. ML model needs to be trained on a set of data where a set of software features are considered as input and the corresponding test cases as output. Finally, the model uses training data to generate new test cases [17].

Test Case Prioritization using ML: Test case prioritization can be performed using machine learning. ML algorithms determine the most critical test cases to execute based on the likelihood of failure and the potential effect on the system. For prioritizing test cases, a machine learning model needs to be trained on a set of labeled data, where a set of software features are considered as input and the corresponding priority level of each test case as output. Finally, the model uses this training data to prioritize new test cases based on their predicted priority level [1].

IV. Systematic Review

A review is a systematic study that helps to identify the existing work, research question improvement scope, and existing empirical studies [18]. In this study, 20 research papers have been reviewed from the past 7 years which have been collected from 6 different databases such as ScienceDirect, IEEE, SCITEPRESS, ACM, Wiley Online Library and MDPI.

A. Eligibility Criteria

Eligibility criteria for selecting articles for a systematic literature review include relevance to the research questions, publication time frame, language, publisher, and study design [19]. For this systematic review, after filtering we have selected 20 articles out of 40 articles from the last 7 years. We only selected articles relevant to software testing using machine learning.

B. Search String Strategy

After lots of searching in the research databases and google scholar, we found many articles about software testing using machine learning but only the most relevant articles were selected. In this process, we have used the advanced search string strategy [20]. In this search string strategy, Boolean operators (AND, OR, NOT) have been used to combine and exclude keywords in the search query.

Our search string was *[("Software Testing" AND "Machine Learning") AND ("Testing Automation Technique" OR "Deep Learning" OR "Black-box Testing" OR "Integration Testing" OR "Metamorphic Testing" OR "White Box Testing") NOT ("Manual Testing" OR "Adhoc testing")].*

Apart from the search string strategy, one can use titles, keywords, or abstracts to find out relevant publications. The purpose of this study is to review the effective application of AI (Machine Learning, Deep Learning) in software testing

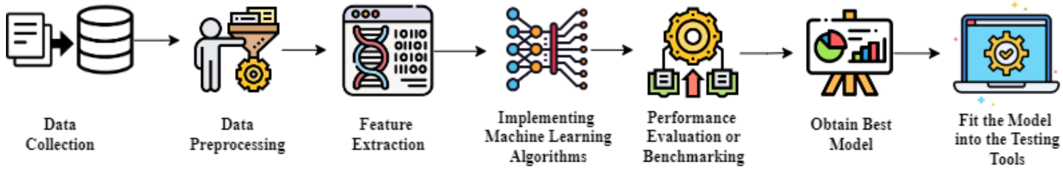

Fig. 1. A General Approach to Apply ML Techniques in Software Testing

TABLE I
INCLUSION AND EXCLUSION CRITERIA

Area	Criteria for Inclusion	Criteria for Exclusion
Article type	Research article, SLR	Poster, Book
Searched keywords	Software testing, Machine learning, testing automation technique, Test data generation, Blackbox testing, Whitebox testing	Keywords other than ones in "Inclusion criteria"
Interest of area	Software testing, Software Engineering, Artificial Intelligence	Area excluding "Inclusion criteria"
Language	English	Languages except English
Time period	2015 -2022	Before 2015

by developing research questions, collecting and selecting proper relatable studies through filtering methods. By examining the existing literature and answering the research questions, we aim to provide the best current practices for software testing.

C. Data Screening and Analysis

Each paper examines different aspects of applications of ML, DL techniques in software testing. In most studies, the authors compare different ML and DL models based on their performance and identify the best results they could generate from those models based on the subject criteria and expected outcome. For the collection process, we have purposively identified 40 research papers that are related to ML, DL, and software testing by searching keywords in google scholar. We also used a backward snowballing method where we checked the references of the selected papers and identified 20 papers. After the collection of papers, we started the screening process where by reading the title we would be able to differentiate whether the topic being addressed is relatable or not. Table I shows lists of inclusion and exclusion criteria in detail during paper selection for this study.

The paper selection process involved sorting based on eligibility criteria with a focus on the automation of software testing using machine learning and deep learning techniques. After filtering, 20 research papers were selected for the literature review study. The search strategy consisted of 5 stages: identification of the research topic, screening, selection of eligible papers, and final inclusion of research articles.

TABLE II
SELECTED RESEARCH STUDIES ACCORDING TO THE PUBLISHER

Publisher Name	# Research Articles
IEEE	8
ACM	6
MDPI	2
Wiley Online Library	1
Science Direct	2
SCITEPRESS	1

D. Data Extraction

Data extraction means the process of retrieving relevant data from various sources for a specific purpose, such as a literature review [21]. In the context of software testing using machine learning and deep learning, data extraction may involve searching through academic journals, conference proceedings, and other sources to gather information on the latest developments and trends in software testing using ML and DL. This information can then be used to summarise a comprehensive review of the current state of the field, identify gaps in existing knowledge, and provide insights into future directions for research and practice. Table II shows the details of the selected number of studies and their publishers.

V. RESULTS

This section provides insights into state-of-the-art techniques and their effectiveness in improving the quality and efficiency of software testing using machine learning and deep learning. The review aims to provide a comprehensive synopsis of the existing research in this domain by analyzing a number of studies. We have selected 20 studies for the study. The details findings of these selected studies have been presented in table III. We also investigated the answer of the research questions from the relevant research papers.

RQ1: Does manual testing have drawbacks?

Manual testing has several drawbacks. Some of the drawbacks of manual testing are it is time-consuming, it does not cover all possible scenarios and use cases, it is costly, it is susceptible to human errors and it can not reproduce test cases accurately [22]. Machine Learning and Deep Learning can help to overcome the mentioned drawbacks of manual testing. ML and DL can automate the testing process. By leveraging the power of algorithms, more accurate testing can be performed [23].

RQ2: Can integration of AI (ML or DL techniques) in software testing help to overcome the drawbacks of manual testing?

Integration of ML and DL techniques in software testing can help to overcome the drawbacks of manual testing by improving the efficiency, accuracy, and effectiveness of the testing process. ML and DL algorithms can be trained to

TABLE III
SUMMARY OF THE SELECTED STUDIES

SL	Paper Id	Year	Publisher Name	Findings
1	P1 [27]	2022	ACM	Authors proposed an approach utilizing Deep Reinforcement Learning (RL) for automating the exploration of Android apps. Authors developed a tool called ARES along with FATE that integrates with ARES.
2	P2 [28]	2022	MDPI	This paper analyzed ML frameworks in the context of software automation and evaluated the performance of testing tools considering various factors. Accuracy or error rate, scope are important factors to determine the effectiveness of frameworks.
3	P3 [29]	2022	Science Direct	This study investigates the efficacy of machine learning, data mining, and deep learning methodologies in predicting software faults. This investigation reveals that data mining and machine learning techniques are utilized more than deep learning techniques.
4	P4 [30]	2022	ACM	This paper introduces Keeper, a novel testing tool. Keeper adopts a unique approach where it creates pseudo-inverse functions for ML APIs. Keeper significantly enhances branch coverage .
5	P5 [31]	2021	IEEE	This study presents DeepOrder, a regression machine learning model based on deep learning techniques. DeepOrder can prioritize test cases and identify failed test cases when it considers various factors such as test case duration and execution status.
6	P6 [32]	2021	Science Direct	This study investigated reward function and reward strategy within the context of continuous integration (CI) testing. The authors proposed three strategies in terms of the reward strategy. Proposed strategies showed promising results.
7	P7 [5]	2021	IEEE	This paper introduces Deep GUI. Deep GUI utilizes deep learning techniques to create a model of valid GUI interactions, based solely on screenshots of applications.
8	P8 [33]	2021	IEEE	This study finds that most ML libraries lack a high-quality unit test suite. Moreover, the study also discovers recurring trends in the unexamined code throughout the five assessed ML libraries.
9	P9 [34]	2021	IEEE	This study presents a deep learning approach to predict the validity of test inputs for RESTful APIs. The proposed network achieved 97% accuracy for the new APIs.
10	P10 [35]	2019	IEEE	This paper introduces Humanoid, a deep learning approach for generating GUI test inputs by leveraging knowledge gained from human interactions. It learns from traces of interactions generated by humans, enabling the automatic prioritization of test inputs based on their perceived importance to users.
11	P11 [36]	2019	ACM	This study finds equivalent mutants are effective for augmenting data and improving the detection rate of metamorphic relations.
12	P12 [37]	2019	MDPI	This study introduces an enhanced CNN model specifically designed to improve the learning of semantic representations from source-code. This study also showed enhancements of the global pattern capture capability of the models which improve the model's generalization performance.
13	P13 [38]	2019	IEEE	This study used three supervised machine learning algorithms for predicting software bugs. To enhance the accuracy of models, random forest ensemble classifiers have been used. The developed models effectively work for various scenarios.
14	P14 [39]	2019	IEEE	This study finds ML algorithms have predominantly been employed in different areas of software testing. Test case generation, evaluation, test oracle construction, and cost predicton for testing activitires can be performed using ML.
15	P15 [40]	2018	ACM	This study presents an approach for automating the test oracle mechanism in software using machine learning (ML). By incorporating a captured component into the application, historical usage data have been gathered. These data later generate an appropriate oracle.
16	P16 [41]	2018	SCITE PRESS	This paper describes a tool that generates test data for programs. The tool operates by clustering input data from a corpus folder and creating generative models for each cluster. These models are recurrent neural networks.
17	P17 [42]	2018	ACM	This paper introduces a methodology called DaOBML, which offers tool support to enhance the quality of environmental models that generate complex artifacts like images or plots. In this study, among six ML algorithms, ANN shows the best performance.
18	P18 [43]	2017	ACM	This study introduces DeepXplore, an innovative whitebox system designed to systematically test DL systems and detect faulty behaviors. DeepXplore can solve joint optimization problems.
19	P19 [44]	2016	Wiley Online Library	This study, proposed a ML approach that can predict metamorphic relations in software programs. To achieve this, authors utilized a graph-based representation of the program.
20	P20 [45]	2016	IEEE	This study proposed an approach for prioritizing test cases in manual testing. The proposed approach considers black-box metadata, including test case history. SVM Rank ML algorithm is used in this study.

TABLE IV
TESTING ACTIVITIES AUTOMATED BY ML & DL

Software Testing Activity	Total
Test Case Generation	4
Defect Prediction	3
Test Case Prioritization	3
Metamorphic Testing	2
Android Testing	2
Test Case Validation	1
White Box Testing	1

automate repetitive testing tasks, which reduces the required effort for manual testing. This improves the efficiency of the testing process and enables faster testing. ML and DL algorithms can also analyze large amounts of data which help to identify defects in the software system. Identification of the defects improve the accuracy of the testing. Apart from that , ML algorithms can generate test cases using historical data or existing code, optimize the testing by prioritizing test cases [24].

RQ3: What software testing tasks can be automated by AI (ML or DL) ?

Machine Learning and Deep Learning techniques can automate different types of software testing tasks such as test results analysis, test case prioritization, defect prediction, test execution, test case evaluation, test case refinement, testing cost estimation, test oracle construction, identification of metamorphic relations, and test case generation [24]. Table IV shows testing activities that can be automated by machine learning

RQ4: What techniques do researchers use to assess AI (ML or DL) when used in software testing?

Researchers consider different performance matrices to assess ML algorithms when used in software testing. The performance matrices are cross-validation, accuracy, precision, recall, receiver operating characteristic (ROC) curve, area under the curve (AUC), and f1 score [25]. The column total represents the number of papers where these testing activities have been automated by machine learning.

Precision: Precision is a statistical measure that quantifies the ratio of true positive instances out of the total positive predictions made. [26].

Recall: Recall is a statistical indicator utilized to quantify the fraction of true positive outcomes within the entirety of actual positive instances [26].

ML and DL algorithms have shown promising results to automate software testing tasks. Some of the promising algorithms are Neural networks, Decision Tree, Support vector machines, and Random Forest, .

VI. CONCLUSIONS

Software testing plays a key role in the development of software. However, as software systems become more complex, traditional manual testing methods are becoming less practical. There has been growing interest in leveraging AI for software testing. The aim of this study is to comprehensively explore the current state of AI in software testing. This review examines various approaches, techniques, and tools employed in this field, assessing their effectiveness. The articles selected for this study were obtained from different research databases using an advanced search strategy. Initially, 40 articles were retrieved, and after a rigorous filtering process, 20 articles were chosen for analysis.

Based on the findings of the selected papers, it is evident that AI (Machine Learning and Deep Learning) can successfully automate several testing tasks. These tasks include Test Case Generation, Defect Prediction, Test Case Prioritization, Metamorphic Testing, Android Testing, Test Case Validation, and White Box Testing. The integration of AI in software testing is shown to simplify testing activities and enhance performance. In the future, incorporating AI (machine learning and deep learning) techniques in different testing activities will make it easier to perform testing activities. A limited number of studies have been examined in this study, which is a limitation. Conducting a review of a larger number of studies would provide the opportunity to gain more deeper insights.

REFERENCES

[1] Sharif, A., Marijan, D. and Liaaen, M., 2021, September. DeepOrder: Deep learning for test case prioritization in continuous integration testing. In 2021 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 525-534). IEEE.

[2] Yang, Y., Li, Z., He, L. and Zhao, R., 2020. A systematic study of reward for reinforcement learning based continuous integration testing. Journal of Systems and Software, 170, p.110787.

[3] Romdhana, A., Merlo, A., Ceccato, M. and Tonella, P., 2022. Deep reinforcement learning for black-box testing of android apps. ACM Transactions on Software Engineering and Methodology (TOSEM), 31(4), pp.1-29.

[4] Li, Y., Yang, Z., Guo, Y. and Chen, X., 2019, November. Humanoid: A deep learning-based approach to automated black-box android app testing. In 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE) (pp. 1070-1073). IEEE.

[5] YazdaniBanafsheDaragh, F. and Malek, S., 2021, November. Deep GUI: black-box GUI input generation with deep learning. In 2021 36th IEEE/ACM International Conference on Automated Software Engineering (ASE) (pp. 905-916). IEEE.

[6] Kanewala, U., Bieman, J.M. and Ben-Hur, A., 2016. Predicting metamorphic relations for testing scientific software: a machine learning approach using graph kernels. Software testing, verification and reliability, 26(3), pp.245-269.

[7] Braga, R., Neto, P.S., Rabêlo, R., Santiago, J. and Souza, M., 2018, September. A machine learning approach to generate test oracles. In Proceedings of the XXXII Brazilian Symposium on Software Engineering (pp. 142-151).

[8] Nair, A., Meinke, K. and Eldh, S., 2019, August. Leveraging mutants for automatic prediction of metamorphic relations using machine learning. In Proceedings of the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation (pp. 1-6).

[9] Singh, P.D. and Chug, A., 2017, January. Software defect prediction analysis using machine learning algorithms. In 2017 7th International Conference on Cloud Computing, Data Science Engineering-Confluence (pp. 775-781). IEEE.

[10] Pan, C., Lu, M., Xu, B. and Gao, H., 2019. An improved CNN model for within-project software defect prediction. Applied Sciences, 9(10), p.2138.

[11] Myers, G.J., Sandler, C. and Badgett, T., 2011. The art of software testing. John Wiley Sons.

[12] Tan, L., Liu, C., Li, Z., Wang, X., Zhou, Y. and Zhai, C., 2014. Bug characteristics in open source software. Empirical software engineering, 19, pp.1665-1705.

[13] Jamil, M.A., Arif, M., Abubakar, N.S.A. and Ahmad, A., 2016, November. Software testing techniques: A literature review. In 2016 6th international conference on information and communication technology for the Muslim world (ICT4M) (pp. 177-182). IEEE.

[14] Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science, 349(6245), pp.255-260.

[15] Efendioglu, M., Sen, A. and Koroglu, Y., 2018. Bug prediction of systemc models using machine learning. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 38(3), pp.419-429.

[16] Hammouri, A., Hammad, M., Alnabhan, M. and Alsarayrah, F., 2018. Software bug prediction using machine learning approach. International journal of advanced computer science and applications, 9(2).
[17] Zhang, D., 2006, November. Machine learning in value-based software test data generation. In 2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06) (pp. 732-736). IEEE.
[18] Kitchenham, B. and Brereton, P., 2013. A systematic review of systematic review process research in software engineering. Information and software technology, 55(12), pp.2049-2075.
[19] Papaioannou, D., Sutton, A. and Booth, A., 2016. Systematic approaches to a successful literature review. Systematic approaches to a successful literature review, pp.1-336.
[20] Mohamed Shaffril, H.A., Samsuddin, S.F. and Abu Samah, A., 2021. The ABC of systematic literature review: the basic methodological guidance for beginners. Quality & Quantity, 55, pp.1319-1346.
[21] Jonnalagadda, S.R., Goyal, P. and Huffman, M.D., 2015. Automating data extraction in systematic reviews: a systematic review. Systematic reviews, 4(1), pp.1-16.
[22] Leitner, A., Ciupa, I., Meyer, B. and Howard, M., 2007, January. Reconciling manual and automated testing: The autotest experience. In 2007 40th Annual Hawaii International Conference on System Sciences (HICSS'07) (pp. 261a-261a). IEEE.
[23] Zhang, J.M., Harman, M., Ma, L. and Liu, Y., 2020. Machine learning testing: Survey, landscapes and horizons. IEEE Transactions on Software Engineering, 48(1), pp.1-36.
[24] Durelli, V.H., Durelli, R.S., Borges, S.S., Endo, A.T., Eler, M.M., Dias, D.R. and Guimarães, M.P., 2019. Machine learning applied to software testing: A systematic mapping study. IEEE Transactions on Reliability, 68(3), pp.1189-1212.
[25] Song, Q., Guo, Y. and Shepperd, M., 2018. A comprehensive investigation of the role of imbalanced learning for software defect prediction. IEEE Transactions on Software Engineering, 45(12), pp.1253-1269.
[26] Tatbul, N., Lee, T.J., Zdonik, S., Alam, M. and Gottschlich, J., 2018. Precision and recall for time series. Advances in neural information processing systems, 31.
[27] Romdhana, A., Merlo, A., Ceccato, M. and Tonella, P., 2022. Deep reinforcement learning for black-box testing of android apps. ACM Transactions on Software Engineering and Methodology (TOSEM), 31(4), pp.1-29.
[28] Fatima, S., Mansoor, B., Ovais, L., Sadruddin, S.A. and Hashmi, S.A., 2022. Automated Testing with Machine Learning Frameworks: A Critical Analysis. Engineering Proceedings, 20(1), p.12.
[29] Batool, I. and Khan, T.A., 2022. Software fault prediction using data mining, machine learning and deep learning techniques: A systematic literature review. Computers and Electrical Engineering, 100, p.107886.
[30] Wan, C., Liu, S., Xie, S., Liu, Y., Hoffmann, H., Maire, M. and Lu, S., 2022, May. Automated testing of software that uses machine learning apis. In Proceedings of the 44th International Conference on Software Engineering (pp. 212-224).
[31] Sharif, A., Marijan, D. and Liaaen, M., 2021, September. DeepOrder: Deep learning for test case prioritization in continuous integration testing. In 2021 IEEE International Conference on Software Maintenance and Evolution (ICSME) (pp. 525-534). IEEE.
[32] Yang, Y., Li, Z., He, L. and Zhao, R., 2020. A systematic study of reward for reinforcement learning based continuous integration testing. Journal of Systems and Software, 170, p.110787.
[33] Wang, S., Shrestha, N., Subburaman, A.K., Wang, J., Wei, M. and Nagappan, N., 2021, May. Automatic unit test generation for machine learning libraries: How far are we?. In 2021 IEEE/ACM 43rd International Conference on Software Engineering (ICSE) (pp. 1548-1560). IEEE.
[34] Mirabella, A.G., Martin-Lopez, A., Segura, S., Valencia-Cabrera, L. and Ruiz-Cortés, A., 2021, June. Deep learning-based prediction of test input validity for RESTful APIs. In 2021 IEEE/ACM Third International Workshop on Deep Learning for Testing and Testing for Deep Learning (DeepTest) (pp. 9-16). IEEE.
[35] Li, Y., Yang, Z., Guo, Y. and Chen, X., 2019, November. Humanoid: A deep learning-based approach to automated black-box android app testing. In 2019 34th IEEE/ACM International Conference on Automated Software Engineering (ASE) (pp. 1070-1073). IEEE.
[36] Nair, A., Meinke, K. and Eldh, S., 2019, August. Leveraging mutants for automatic prediction of metamorphic relations using machine learning. In Proceedings of the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation (pp. 1-6).
[37] Pan, C., Lu, M., Xu, B. and Gao, H., 2019. An improved CNN model for within-project software defect prediction. Applied Sciences, 9(10), p.2138.
[38] Immaculate, S.D., Begam, M.F. and Floramary, M., 2019, March. Software bug prediction using supervised machine learning algorithms. In 2019 International conference on data science and communication (IconDSC) (pp. 1-7). IEEE.
[39] Durelli, V.H., Durelli, R.S., Borges, S.S., Endo, A.T., Eler, M.M., Dias, D.R. and Guimarães, M.P., 2019. Machine learning applied to software testing: A systematic mapping study. IEEE Transactions on Reliability, 68(3), pp.1189-1212.
[40] Braga, R., Neto, P.S., Rabêlo, R., Santiago, J. and Souza, M., 2018, September. A machine learning approach to generate test oracles. In Proceedings of the XXXII Brazilian Symposium on Software Engineering (pp. 142-151).
[41] Paduraru, C. and Melemciuc, M.C., 2018, July. An Automatic Test Data Generation Tool using Machine Learning. In ICSOFT (pp. 506-515).
[42] de Santiago, V.A., da Silva, L.A.R. and de Andrade Neto, P.R., 2018, September. Testing environmental models supported by machine learning. In Proceedings of the III Brazilian Symposium on Systematic and Automated Software Testing (pp. 3-12).
[43] Pei, K., Cao, Y., Yang, J. and Jana, S., 2017, October. Deepxplore: Automated whitebox testing of deep learning systems. In proceedings of the 26th Symposium on Operating Systems Principles (pp. 1-18).
[44] Kanewala, U., Bieman, J.M. and Ben-Hur, A., 2016. Predicting metamorphic relations for testing scientific software: a machine learning approach using graph kernels. Software testing, verification and reliability, 26(3), pp.245-269.
[45] Lachmann, R., Schulze, S., Nieke, M., Seidl, C. and Schaefer, I., 2016, December. System-level test case prioritization using machine learning. In 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 361-368). IEEE.