Using machine learning to speed up automated testing

Kehinde Esther Ogundeyi

*Institute of Computer Science University of Tartu*

Tartu, Estonia [kehinde.esther.ogundeyi@ut.ee](mailto:kehinde.esther.ogundeyi@ut.ee)

Sammar Ahmad

*Institute of Computer Science University of Tartu*

Tartu, Estonia [sammar.ahmad@ut.ee](mailto:sammar.ahmad@ut.ee)

***Abstract*—Test Automation helps to manage test data, perform tests, and evaluate test results to enhance the software quality of a software application. There are known issues with automation testing, such as slowness, flakiness, and test maintainability. This paper is a systematic mapping study to give an overview of how Machine Learning can solve the above problems.**

***Keywords—Test Automation, Machine Learning, Speeding Up Automated Testing***

1. INTRODUCTION

Machine learning has made significant advances in a variety of domains in the industry. Machines began to comprehend verbal commands, analyse data, identify images, and drive automobiles because of the incredible progress expanded in the use of technology and computer hardware [1]. Artificial Intelligence began when computers were fast and had a large memory playing an essential part in a variety of sectors, one of which is that software testing is a must-do procedure [2]. It ensures the fulfilment of business requirements and leads to consumer pleasure, as well as a successful voyage across the Lifecycle of software development. We shall review machine learning techniques used in automated testing in this paper, which have significantly improved testing.

1. METHODOLOGY

The following five steps are used to complete the supplied article, which is a systematic mapping study: (1) research questions are raised, (2) data is obtained, (3) high-quality studies are chosen, (4) data is retrieved, and (5) conclusions are drawn.

1. *Research Questions*

The goal of this systematic mapping research methodology is to locate relevant literature on the topic of machine learning in speeding up automated testing. The main aim is to gain an overview of how machine learning can help in speeding up the automated testing process, and what possible machine learning techniques are available to achieve this goal. Four research questions (RQs) were proposed to achieve this goal:

* + RQ1: What are the advantages of machine learning in automated testing?
  + RQ2: What are the available machine learning techniques for automated testing?
  + RQ3: What are the impacts of current machine learning techniques for automated testing?
  + RQ4: Which machine learning technique will be the most effective in speeding up automated tests?

1. *Data Search*

The Database IEEE Xplore and springer Link were selected for this study. As they have a huge number of computer science and engineering articles on the internet. This search string was created to search this data source:

(Test automation) AND ( machine learning) AND ( machine learning techniques) AND (Software testing Areas)

Fig. 1. The search string

The reason for using this platform was that there were very few publications available related to the topic. To exclude other dimensions of machine learning, the search query included the advantages of machine learning in automated testing. The query returned 40 results with no additional filtering.

1. *Study Selection and Quality Assessment*

We searched through IEEExplore to search for publications related to our topic. The following exclusion criteria were created:

* + EC1: Removed papers published than 2010 (23 remaining)
  + EC2: Papers published before 2016 (28 remaining)
  + EC3: Papers that were not in English (17 remaining)
  + EC4: Papers that were inaccessible (15 remaining)

A set of inclusion criteria was created:

* + IC1: Include articles that are about automated testing and have machine learning in it
  + IC2: Articles discussing the various advantages of machine learning
  + IC3: Articles that discussed the various machine learning techniques used in automated testing

EC1 was applied to the initial set of 40 publications, yielding 23 results. EC2 was used to filter out publications that are not articles, yielding 20 remaining papers. Only 15 Papers remained after applying the exclusion criteria. Each of the 15 publications' abstracts was personally examined after exclusion criteria were applied. The abstracts were analysed to determine the inclusion criteria. As a result, a total of 7 publications were included in this stud

1. *Data Extraction*

A data extraction table, which comprises several data items, was utilized for data extraction (Table 1). The article link and title, author name(s), services, fields and domains, technologies and devices, and future solutions. The papers collected for this study were read and analysed to see what data they contained, such as whether they mentioned any machine learning or automated testing, what fields or domains these services are used in, whether they contained information about machine learning techniques used in automated testing, or whether they discussed possible new methods to implement.

TABLE I. DATA EXTRACTION FORM

|  |  |
| --- | --- |
| **Advantages** | **Occurrences** |
| Better Automation | 7 |
| Reusability | 8 |

Table II shows the occurrences of different advantages that can be achieved when we use Machine Learning with Automated Testing and it also gives the advantages of using machine learning in automated testing; among them, the most commonly mentioned were improved performance of coverage testing, removing redundant tests, data generation and making accurate predictions. In addition, prioritizing testing appropriately, increasing productivity, and Better Test Code Maintenance was also considered high among other papers. Prioritizing testing appropriately and Focusing on unique features were mentioned as advantages benefited in using machine learning.

|  |  |  |
| --- | --- | --- |
| **Data Item** | **Value** | **RQ** |
| Article link | IEEEXplore link to article |  |
| Article title | Name of the article |  |
| Author  name(s) | Set of Names of the  authors |  |
| Advantages | Advantages of machine learning in automated testing | RQ1 |
| Available techniques | Currently Available machine learning techniques | RQ2 |
| Impacts | Impacts of using currently available techniques | RQ3 |
| Most Effective | Which machine learning technique will be the most effective in speeding up automated tests? | RQ4 |

There are other mentions for ensuring quality assurance, helping both testers and developers, improve software quality with better automation by expanding the depth and scope of testing, and Reusability.

*B) RQ2:* What are the available machine learning techniques for automated testing*?*

TABLE III. Machine Learning Techniques

|  |  |
| --- | --- |
| Available Machine Learning techniques | **Occurrences** |
| Decision Tree | 6 |
| Linear regression | 7 |
| AdaBoost | 8 |
| Random Forest | 5 |
| Hybrid genetic algorithm (HGA) | 4 |
| Support Vector Regression (SVR) | 4 |
| K-Means Clustering | 7 |
| Genetic Algorithm | 5 |

1. RESULTS

The articles' responses were compiled into tables that represented each research question.

*A) RQ1:* *What are the advantages of machine learning in automated testing?*

TABLE II. ADVANTAGES

Table III presents us with all the available techniques that have been used in automated testing. Decision Tree which is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems, Linear regression is a basic and commonly used type of predictive analysis, AdaBoost short for Adaptive Boosting, is a Boosting technique used as an Ensemble Method in Machine Learning, and Random Forest which is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems have the most references.

|  |  |
| --- | --- |
| **Advantages** | **Occurrences** |
| Early Indications of Problems/Defects | 9 |
| Prioritize Testing Appropriately | 6 |
| Improve Performance | 10 |
| Make accurate predictions | 10 |
| Focus on unique features | 8 |
| Increase productivity | 9 |
| Ensures quality assurance | 4 |
| Help both testers and developers | 5 |
| Better Test Code Maintenance | 8 |

Other techniques are also Hybrid genetic algorithm which uses a combination of possible solution to get a solution, Support Vector Regression is widely used for classification problems in machine learning, K-means clustering is one of the simplest and popular unsupervised machine learning algorithms, it aims to partition n observations into k clusters and Genetic Algorithm has the concepts of genetics that looks to at the population of individual crossover which improves iteration by iteration producing a better result

1. *RQ3: What are the impacts of current machine learning techniques for automated testing?*

Different Machine Learning models have impacted or solved different sets of test automation problems. The below table can explain what models are used for which problems

TABLE IV. IMPACTS

|  |  |  |
| --- | --- | --- |
| **Software Testing Areas** | **Machine Learning techniques** | **Impacts** |
| Fault localisation | Decision Tree | Fault localization indicates code sections that are likely  to need to be changed to  address known flaws and  leveraging on this has  significant effect in  enhancing defect prediction |
| coverage prediction results | Support Vector Regression (SVR) | The use of support vector regression in coverage predictions is that it improves testing  quality, produce smarter and more  accurate test cases for systems. |
| Test case Classification | K-Means Clustering | Using k-means clustering for test case classification methodology has have a god impact by enhancing regression testing |
| Automatically Testing GUI | K-Means Clustering | Automatically testing GUI with K-Means clustering has be proven to improve test sequence, test case optimization, as well as identifying infeasible GUI test cases. |
| Detection of codes that might need improvement | Adaboost and Random Forest | As a result of this technique,  they have proposed a reduction  in the efforts that are put in  the testing of software  because classes that are prone to change during the early stages of development be discovered earlier. |
| Test data generation | Genetic Algorithm | Software testing costs can be cut by using automated test data generation to discover the smallest collection of data with the most coverage. choosing to complete the most important testing objectives, developing a more cost-effective strategy to complete the remaining testing targets, and achieving a graceful deterioration when the testing budget is cut |
| Test Failure Prediction | Linear Regression | The technique allowed for more efficient Regression Testing by detecting failures early. |
| Performance Analysis | AdaBoost | The research done by [8] examined the performance of refactoring decision tool in reducing class maintainability and found that using AdaBoost considerably aids the software team in the evaluation process. |

All references in the papers to any technologies or devices, which are used to implement location-tracking, are shown in Table 4.

In the context of software testing, machine learning has a wide range of techniques and algorithms. The algorithms and techniques used in Machine Learning differ in terms of how they function, their mathematical and statistical models, assumptions, characteristics, accuracy, strengths and weaknesses, and the problem category they solve if they solve classification, regression, or other issues and in Table 4, we showed the different software testing areas in which these techniques have been applied.

1. *RQ4:* *Which machine learning technique will be the most effective in speeding up automated tests?*

We considered AdaBoost deemed the best approach for handling the majority of test automation challenges because The AdaBoost technique has the potential to increase prediction accuracy [6]. From the experiment done in [8], Adaboost has the highest accuracy score with 99.6%. In addition, in [9] Models for forecasting the change likeliness of feature systems were created in research. The generated models could be used to anticipate alter classes during the software development process. With a score of 0.877, Adaboost has the best accuracy. Other algorithms, such as random forest and bagging, have shown to be competitive.

1. DISCUSSION

*A) RQ1:* *What are the advantages of machine learning in automated testing?*

The Advantages shown in table II which mentioned the benefits of machine learning in automation. The results are plausible and explainable because comparing research done by [1], [3], and [4] there was more similarities in improving performance as well as better source code maintenance, Other research done by [5], [6] and [7] also compliment the advantage pf having improved performance but also emphasis on making accurate predictions. With more research being done more advantages can still found when machine learning is applied to automated testing as this finding could be explored further to investigate more advantages in term of making accurate predictions and quality assurance.

The limitations of the survey study we encountered some of artificial intelligence advantages related to artificial intelligence.

*B) RQ2:* What are the available machine learning techniques for automated testing*?*

From the research we did, different machine learning techniques were mentioned in [3], [6] and [8]. The result is

plausible and also explainable but there were some surprising findings as some papers such as [6] applied both techniques of Machine learning and Natural learning process to improve some testing areas like detecting defect duplicate report. There was no contradiction between the papers but rather they complement each as there were common mention of the machine learning techniques among the papers we researched. More findings can still have done as there are many other machine learning techniques that can be applied to automated testing. The limitations of the survey study is that we encountered some of artificial intelligence techniques applications to automated testing rather than machine learning

*C) RQ3: What are the impacts of current machine learning techniques for automated testing?*

From our research, we were able to find out machine learning techniques and how they are applied to various testing area in automated testing. We compare the results of different techniques such as to improve regression testing, the authors suggested a test case classification methodology based on k-means clustering. This research also discovered that using the statement coverage requirement first improves the performance of the clustering-based technique based on article [8]. There were some interesting outcomes, since several studies, such as [7], used both machine learning and natural learning techniques to improve specific testing areas, such as defect duplicate report detection. The method we used to gather these techniques is that we compare a series of experiments done by different articles which used publicly accessible annotated dataset of diverse software systems to see how well their proposed technique performs. This research can help further research in determining the impacts in which the various machine learning techniques have impacted automated testing, further research can also still be done to determine other testing areas not mentioned in this paper.

*D) RQ4:* *Which machine learning technique will be the most effective in speeding up automated tests?*

From our research, we looked at different aspects in which the machine learning techniques were applied and try to compare their accuracy, the result was related as AdaBoost have better performance in the papers we research, though they have applied it in different testing areas, AdaBoost still have some good performance compared to other techniques listed in table III were applied in different sectors with different results. Though there were some limitations in aspect of doing the comparison as all papers did not use exactly the same techniques for their research. This finding can still be explored by reviewing other machine learning techniques not mention and compare them with Adaboost result to view how it performs when compared

CONCLUSION

For this paper, a systematic mapping study was carried out. Research questions were answered concerning the four research questions about machine learning and automated testing. The study focused mainly on machine learning techniques used in automated testing; most of the found data covered these fields. Many papers discussed the advantages of machine learning in automated testing(RQ1), like performance improvement and accurate predictions. (RQ2) describes some of the available techniques, and the most common are decision tree, K nearest neighbours, and Random forest. In (RQ3), we grouped the techniques according to their software testing areas which we discussed the best in (RQ4).

1. REFERENCES
2. Laurie Butgereit “Using Machine Learning to Prioritize Automated Testing in an Agile Environment” 2018 IEEE International Conference on Software Testing, Verification and Validation Workshops
3. Jerry Gao, Chuanqi Tao, Dou Jie, Shenqiang Lu “What is AI Software Testing? and Why” 2019 IEEE International Conference on Service-Oriented System Engineering (SOSE).
4. Ariel Rosenfeld, Odaya Kardashov, and Orel Zang “Automation of Android Applications Testing Using Machine Learning Activities Classification” arXiv:1709.00928v1 [cs.SE] 4 Sep 2017
5. Adam Tornhill “Assessing technical dept in automated tests with codeScene” 2018 IEEE International Conference on Software Testing, Verification and Validation Workshops.
6. Vinod Yadav1, Raphael Kwaku Botchway,Roman Senkerik, Zuzana Oplatkova Kominkova4 “Robot Testing from a machine learning perspective” Proc. of the International Conference on Electrical, Computer and Energy Technologies (ICECET) 9-10 December 2021, Cape Town-South Africa
7. Hussam Hourani, Ahmad Hammad, Mohammad Lafi “The Impact of Artificial Intelligence on Software Testing” 2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)
8. Vinicius H. S. Durelli , Rafael S. Durelli , Simone S. Borges, Andre T. Endo , Marcelo M. Eler ,Diego R. C. Dias , and Marcelo P. Guimar˜aes “Machine Learning Applied to Software Testing: A Systematic Mapping Study” IEEE TRANSACTIONS ON RELIABILITY, VOL. 68, NO. 3, SEPTEMBER 2019

[8] Lov Kumar,Shashank Mouli Satapathy,Ashish Sureka “”

[9] A. K. Tripathi and K. Sharma, “Optimizing testing efforts based on

change proneness through machine learning techniques,” 2014 6th

IEEE Power India International Conference (PIICON), 2014.

[10] Leveraging Fault Localisation to Enhance Defect Prediction

[11] Machine Learning-Based Radio Coverage Prediction in Urban Environments

[12] Automated test data generation for branch testing using incremental genetic algorithm

[13] https://ieeexplore.ieee.org/document/4031966