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 Automation 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 automation testing. The main aim is to gain an overview of how machine learning can help in speeding up the automation 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 over traditional 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 was 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:

("All Metadata": test automation) AND ("All Metadata": machine learning) AND ("All Metadata": speeding up)

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 automation 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 automation 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 automation 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 study.

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 7 papers collected for this study were read and analysed to see what data they contained, such as whether they mentioned any machine learning or automation testing, what fields or domains these services are used in, whether they contained information about machine learning techniques used in automation testing, or whether they discussed possible new methods to implement.

TABLE I. DATA EXTRACTION FORM

|  |  |
| --- | --- |
| **Advantages** | **Occurrences** |
| Better Automation | 1 |
| Reusability | 1 |

Table II shows the occurrences of different advantages that can be achieved when we use ML with Automation Testing.

Table II gives the advantages of using machine learning in automation testing; among them, the most commonly mentioned were improved performance 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 twice.

|  |  |  |
| --- | --- | --- |
| **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 automation testing | RQ1 |
| Available techniques | Currently Available machine learning techniques | RQ2 |
| Impacts | Impacts of using currently available techniques | RQ3 |
| Most Effective | Best Machine learning techniques in automation testing | RQ4 |

There are single mentions for ensuring quality assurance, Helping both testers and developers, Better Automation, and Reusability.

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

TABLE III. TECHNIQUES

|  |  |
| --- | --- |
| Available techniques | **Occurrences** |
| Decision Tree | 2 |
| Linear regression | 2 |
| AdaBoost | 2 |
| Random Forest | 2 |
| Hybrid genetic algorithm (HGA) | 1 |
| Support Vector Regression (SVR) | 1 |
| K-Means Clustering | 1 |
| Genetic Algorithm | 1 |
| Natural Language Processing (NLP) | 3 |

1. RESULTS

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

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

TABLE II. ADVANTAGES

Table III presents us with all the available techniques that have been used in automated testing. Decision Tree, Linear regression, AdaBoost, NLP, and Random Forest have the most references, while the other techniques have been mentioned once in the publications

|  |  |
| --- | --- |
| **Advantages** | **Occurrences** |
| Early Indications of Problems/Defects | 3 |
| Prioritize Testing Appropriately | 2 |
| Improve Performance | 4 |
| Make accurate predictions | 4 |
| Focus on unique features | 2 |
| Increase productivity | 3 |
| Ensures quality assurance | 1 |
| Help both testers and developers | 1 |
| Better Test Code Maintenance | 3 |

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** | **Algorithm** |
| Fault Detection | Decision Tree |
| coverage prediction results | Support Vector Regression (SVR) |
| Automatically Testing GUI | K-Means Clustering |
| Test case Classification | K-Means Clustering |

|  |  |
| --- | --- |
| **Software Testing Areas** | **Algorithm** |
| Detection of duplicate defect report | Natural Language Processing (NLP) |
| Change Proneness of Object Oriented | Adaboost and Random Forest |
| Test data generation | Genetic |
| Test Case Prioritization | Natural Language Processing (NLP) |
| Early Detection of Bugs | Natural Language Processing (NLP) |
| Test Failure Prediction | Linear Regression and Natural Language Processing (NLP) |
| Automatically generation of test case | Natural Language Processing (NLP) |

All references in the 16 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 AI 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 achieving this objective?*

Based on the results of RQ3, Natural Language Processing was deemed the best approach for handling the majority of test automation challenges.

We considered Natural Learning processing to be the most effective because it was used in most software testing areas such as; Test case Prioritization, Early detection of Bugs, Test failure prediction, Automatically generation of the test case, and Detection of duplicate defect report and part of this result is that it reduces the amount of time and effort required by software testers to test the product, increase the testing efficiency of software and in addition, The NLP-based technique has the potential to increase prediction accuracy [6].

1. DISCUSSION

CONCLUSION

For this paper, a systematic mapping study was carried out. Research questions were answered concerning the four research questions about machine learning and automation 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