



A Machine Learning Approach for Context-Aware Input/Output Validation in Mobile Applications

H.K.B De Dilva
(Reg. No.: MS23002388)

A THESIS
SUBMITTED TO
SRI LANKA INSTITUTE OF INFORMATION TECHNOLOGY
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE IN INFORMATION TECHNOLOGY

December 2024

I certify that I have read this thesis and that in my opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Science.

Mr. Thusithanjana Thilakarathna

Approved for MSc. Research Project:

MSc. Programme Co-ordinator, SLIIT

Approved for MSc:

Head of Graduate Studies, FoC, SLIIT

DECLARATION

This is to certify that the work is entirely my own and not of any other person, unless explicitly acknowledged (including citation of published and unpublished sources). The work has not previously been submitted in any form to the Sri Lanka Institute of Information Technology or to any other institution for assessment for any other purpose.

Sign: 

H.K.B De Silva

Date: 14/09/2024

ABSTRACT

A Machine Learning Approach for Context-Aware Input/Output Validation in Mobile Applications

Kushantha De Silva

MSc. in Information Technology

Supervisor: Mr. Thusithanjana Thilakarathna

December 2024

There is still insecurity in mobile application since, input/output validation is not well implemented since the rule-based methods cannot adapt the new attacking forms and the new environments. Thus, this work puts forward a novel approach for context-aware input/output validation in mobile applications to overcome these challenges with the use of machine learning. The work is targeted towards investigating a sequence of previous data, application context, and user input for identifying abnormal patterns in real-time using a machine learning model. In line with the formulated model, an adaptive validation system will be employed so that the validation criteria are fluid with the detected context and possible threats. To measure the impact and satisfaction level of the proposed system, this study will use both penetration test and users. Penetration testing will establish the effectiveness of the model in discovering and even preventing security threats while user research will determine the ease of use of the application with the implemented security method. The hope is that this research will be of significant value in formulating mobile applications that are more secure while at the same time providing users with a positive experience. In conclusion, this machine learning integrated method of validation seeks to enhance application security as well as satisfaction levels of the users.

Keywords: mobile applications, input/output validation, machine learning, anomaly detection, context-aware, user experience

ACKNOWLEDGEMENT

Firstly, I would like to express my thankfulness to my supervisor Mr. Thusithanjana Thilakarathna (Lecturer – Faculty of Computing, SLIIT) for his great support towards the research and Almighty God for giving success in providing me the right path for the improvement for the successful completion of the research. From him I was taught the right way to enter competitions, to collaborate with professionals, demonstrate my potential and achieve success for SLIIT. Moreover, I would like to express my heartfelt appreciations for Dr. Dilshan De Silva (MSc Programme coordinator for IT) and Dr. Prasanna S. Haddela (MSc Programme coordinator) for the counseling done for me. Also, my sincere thanks to Ms. Gihani Samaranayake (QA Engineer - Antler IT Solutions (Pvt) Ltd) for all her guidance provided during the testing process of the system with regards to her expertise in the stream. And also, another special thanks should be goes to Mr. Nalaka Wimalaratne (CEO of Antler IT Solutions (Pvt) Ltd), Ms. Madhusha Wijetunga (Senior Engineer IT & Support - Antler IT Solutions (Pvt) Ltd) and Mr. Chamal Fernando (Engineer IT & Support - Antler IT Solutions (Pvt) Ltd) who gave me the guidance, permissions and facilities for the Testing phases. I also would like to extend my special thanks to Mr. Gayashan Perera who worked as a software engineer at Pearson Lanka (Pvt) Ltd, and Mr. Deneth Perera who worked as a senior software engineer at Antler IT Solutions (Pvt) Ltd for the effort given to support in packing this final product. Also, I would like to give appreciation to my parents and wife for their patience and for offering their time as well as offering resources to get the necessary and other requirements. Many thanks to other fellow colleagues who have been provided with support for this to make this research more efficient and successful.

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT	iii
ACKNOWLEDGEMENT	iv
TABLE OF CONTENTS	v
List of Figures	vii
List of Tables.....	viii
Chapter 1 Introduction	1
1.1 Background and Literature Review.....	1
1.1.1 Background	1
1.1.2 Literature Review	7
1.2 Research Gap	9
1.3 Research Problem	14
1.4 Research Objectives	15
Chapter 2 Research Methodology	16
2.1 Methodology	16
2.1.1 Describing Methodological Steps.....	16
2.1.2 Research Area	22
2.1.3 Data Collection	22
2.1.4 Design	23
2.1.5 Tools and Technologies	24
2.2 Commercialization Aspects & Cost	25
2.2.1 Commercialization Aspects	25
2.2.2 Cost	26
2.3 Implementation & Testing	26
2.3.1 Implementation	26
2.3.2 Testing Phase 1: QA & User Involved Testing	31
2.3.3 Testing Phase 2: Ensuring Thorough Validation	35
Chapter 3 Results and Discussion	37
3.1 Results	37
3.2 Discussion.....	49
3.2.1 Accuracy Discussion Among the Selected Model with Others	49

3.2.2 Discussion of Testing Phase 1	52
3.2.3 Discussion of Testing Phase 2	59
Chapter 4 Summary of Student Contribution	62
Chapter 5 Conclusion	65
References	69
Appendix	72
Appendix 1: Process of ML Model Training.....	72
Appendix 2: Penetration Testing Report	73
Appendix 3: Sample Input with Resulted Outputs via Postman.....	77
Appendix 4: Screenshots of the Process of API Testing conducted on the Production Server ..	80
Appendix 5: Test Report of API Testing Conducted on the Production Server.....	81
Appendix 6: Screenshots of selected and IEEE Copyright transfer confirmation mails of ICAC 2024	82

List of Figures

Figure 1.1 Percent of Applications with security flaws over past months	1
Figure 1.2 Comparison OWASP mobile top 10 between year 2016 and 2024.....	3
Figure 2.1 Sample SQL Injection Attacks	17
Figure 2.2 Connectivity of Methodological Steps and output components	22
Figure 2.3 Leveling Structure.....	27
Figure 2.4 File Structure of API Implementation.....	30
Figure 3.1 Feature Importance in RandomForest Classifier	40
Figure 3.2 Accuracies and other stats.....	42
Figure 3.3 Confusion Matrix.....	45
Figure 3.4 Scatter Plot of Actual vs Predicted Labels.....	47
Figure 3.5 Classification Reports of all trained model with Accuracies	49

List of Tables

Table 1.1 Comparison between the related works and the proposed approach	10
Table 1.2 Gap discussions to relevant criteria of related works by mentioning the proposed approach	11
Table 2.1 Table of Cost	26
Table 2.2 Applied Strategies for the Model Evaluation	28
Table 2.3 Penetration Testing Aspects Categorization.....	35
Table 2.4 Aspects Categorization of Production Server Testing Process	36
Table 3.1 Summary of Model Comparison	51
Table 3.2 Testing Outcomes	56
Table 3.3 Test Results of API testing on Production Server.....	59
Table 3.4 Comparison between the Penetration and the Production Sever Tests	61