

# Enhancing the Security of Web Sites and Patients' Portals by Detecting Malicious Web Robots Using Machine Learning Techniques

Nafiseh HOSSEINI<sup>1</sup>, Kobra ETMINANI<sup>1</sup>, Farhad FATEHI<sup>2</sup>

1. Mashhad University of Medical Science, Mashhad, Iran
2. The University of Queensland Centre for Online Health, Queensland, Australia

## Introduction

There is increasing demand for access to medical information via web sites and patients' portals, but one of the challenges towards widespread utilization of such service is maintaining the security of those web sites and portals. Recent reports show an alarming increase in cyber-attacks using web robots. These software programs crawl web pages and are capable of executing various commands such as attacking web servers, cracking passwords, harvesting user's personal information, and testing the vulnerability of servers.

## Aim

The aim of this research was to develop a new effective model for detecting malicious web robots using machine-learning techniques.

## Method

In this research, different methods of web robots detection were investigated. Log files of a sample of compromised web sites were analyzed and the best features for the detection of web robots were extracted. Then after testing and comparing several machine learning algorithms including Support Vector Machine (SVM), Bayesian Network and Decision Tree, the best model was developed using the most appropriate features and its accuracy was evaluated.

## Results

Our analyses showed the SVM-based models can yield higher accuracy ( $f=0.97$ ) comparing to Bayesian Network ( $f = 0.88$ ) and Decision Tree ( $f = 0.95$ ) for detecting malicious web robots. However, extracting proper features can increase the performance of the Bayesian network ( $f = 0.94$ ) and the Decision Tree ( $f = 0.96$ ).

## Conclusion

Security concerns are among the potential barriers to widespread utilization of patient portals. Machine learning algorithms can be effectively used to detect malicious web robots and enhance the security of sensitive patients' information. Selecting appropriate features for the development of these algorithms can remarkably increase their accuracy.

## Correspondence:

Nafiseh Hosseini  
Mashhad University of Medical Science, Mashhad, Iran  
Hosseini84.stphd@gmail.com

---

---

---

---

---

---

---

---

---

---