وزارة التعليم العالي والبحث العلمي
جامعة الموصل
كلية علوم الحاسوب والرياضيات
قسم البرمجيات

كشف ثغرات البرمجيات باستخدام التعلم العميق

رسالة مُقدَّمة
إلى مجلس كلية علوم الحاسوب والرياضيات في جامعة الموصل
وهي جزء من متطلبات نيل شهادة ماجستير علوم في
البرمجيات

من قبل
شهد جلال أحمد الطائي

إشراف
أ.د. دجان بشير طه

1445 هـ
2023 م

Ministry of Higher Education and
Scientific Research
University of Mosul
College of Computer Science and
Mathematics
Department of Software

Software Vulnerabilities Detection Using Deep learning

**A Thesis Submitted to the Council of the College of
Computer Science and Mathematics
University of Mosul
as a Partial Fulfillment of Requirements
for the Degree of Master of Science
in
Software**

By
Shahad Jalal Ahmed

Supervised by
Prof. Dr. Dujan Basheer Taha

2023A.D. **1445A.H.**

Abstract

The process of detecting software vulnerabilities is considered a significant challenge for researchers, as they strive to achieve accuracy and excellent performance in this vital field. These challenges arise from the urgent need to discover vulnerabilities in a wide range of domains dealing with digital documents, including software security and protection against intrusions and other security risks. Additionally, technological advancements and the increasing complexity of software continue to make research in this field more difficult, necessitating urgent efforts and investment in research and innovation to keep up with advanced cyber threats. Continuous efforts and investment in research and innovation are required to develop advanced techniques that contribute to the accurate and effective discovery of software vulnerabilities. Researchers need the necessary skills to deal with software complexities and comprehensively test them to identify and fix potential vulnerabilities.

In this study, a hybrid algorithm combining the advantages of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) for reading and classifying sequential data in software functions. This algorithm excels in handling sequentially varying input lengths, allowing it to deal with different dimensions in software functions without the need to specify or indicate those dimensions during the training phase. The Draper database, containing 1.27 million functions extracted from open-source software, was used to train this algorithm. Providing a sufficient amount of data for training and using pre-processing techniques are crucial factors in building a strong and effective model in this field.

The results demonstrated that the proposed model achieves significantly higher accuracy compared to previous researches, with an accuracy rate of 93.53%. The recall rate reached 90%, while the precision rate was 62.07%. The F1-Score, which combines precision and recall, was calculated for the model and reached 73.47% in the process of identifying software vulnerabilities.