# An Enhanced Model of Whale Optimization Algorithm and K-nearest Neighbors for Malware Detection Detection

**Prepared by**

**Mariam Kamal Al-ghamri**

**Supervised by**

**Dr. Rami Sihwail**

## Abstract

Malicious software is a program that is designed to penetrate or damage a computer system without the consent of the owner, and have become a major threat to the security of systems and networks. The exponential growth in data volume and feature dimensionality poses challenges in machine learning (ML) and various fields, resulting in reduced classification accuracy and heightened computational costs. Feature selection is a crucial preprocessing step which addresses these challenges by eliminating irrelevant, redundant, and less informative features that may adversely impact classifier performance. This study introduces an enhanced Whale optimization algorithm (EWOA) by utilizing Opposite-Based Learning (OBL) and proposing a new search mechanism that is aiming to improve classification accuracy, feature selection, and overall malware detection model efficiency. The WOA

serves as a metaheuristic general-purpose algorithm designed for solving continuous search problems. In contrast to the traditional WOA, the proposed EWOA is more capable to avoid falling in local optima and provides an improved search mechanism that includes mutation and neighborhood search strategies to enhance search capabilities. Furthermore, EWOA enhances population diversity using OBL technique. Performance evaluations on the CIC-MalMem-2022 dataset were conducted. In order to confirm the effectiveness of the proposed method, we compared the average number of features, efficiency, fitness value, accuracy, and statistical tests among the following optimization algorithms: Gray Wolf optimization algorithm (GOA), Genetic optimization algorithm (GA), Particle swarm optimization (PSO), Artificial Lion optimization (ALO), Butterfly optimization algorithm (BOA), and Slime Mould algorithm (SMA). The experimental results have confirmed the dominance of EWOA over the other optimization algorithms, based on 30 runs, in different aspects, such as accuracy (0.99987%), fitness value (0.00084511%), and feature selection (average of 3.97 features).

# النموذج المحسن لخوارزمية الحوت التحسينية وخوارزمية الجار القريب لكشف البرمجيات الخبيثة

**إعداد**

**مريم كمال الغمري**

**إشراف**

**الدكتور رامي سحويل**

## الملخص

البرامج الضارة هي برامج مصممة لاختراق نظام الكمبيوتر أو إتلافه دون موافقة المالك، وأصبحت تشكل تهديدًا كبيرًا لأمن الأنظمة والشبكات .يفرض النمو الهائل في حجم البيانات وأبعاد الميزات تحديات في التعلم الآلي ومختلف المجالات، مما يؤدي إلى انخفاض دقة التصنيف وزيادة التكاليف الحسابية .يعد اختيار الميزة خطوة حاسمة في المعالجة المسبقة والتي تعالج هذه التحديات من خلال إزالة الميزات غير ذات الصلة والمتكررة والأقل إفادة، والتي قد تؤثر سلبًا على أداء المصنف .تقدم هذه الدراسة خوارزمية تحسين الحوت المحسنة (EWOA) من خلال استخدام التعلم القائم على العكس (OBL) واقتراح آلية بحث جديدة تهدف إلى تحسين دقة التصنيف واختيار الميزات وكفاءة نموذج الكشف عن البرامج الضارة بشكل عام .تعمل خوارزمية تحسين الحيتان بمثابة خوارزمية ذات أغراض عامةmetaheuristic مصممة لحل مشكلات البحث

المستمر .على النقيض من WOA التقليدية، فإن EWOA المقترحة أكثر قدرة على تجنب الوقوع في الحلول المحلية وتوفر آلية بحث محسنة تتضمن استراتيجيات البحث عن الطفرات والجوار لتعزيز قدرات البحث .علاوة على ذلك، تعمل EWOA على تعزيز التنوع في اختيار الحيتان في المرحلة الابتدائية باستخدام تقنية OBL. تم إجراء تقييمات الأداء على مجموعة بيانات CIC-MalMem-2022. من أجل تأكيد فعالية الطريقة المقترحة، قمنا بمقارنة متوسط عدد الميزات والكفاءة وقيمة اللياقة البدنية والدقة والاختبارات الإحصائية بين خوارزميات التحسين التالية: خوارزمية تحسين الذئب الرمادي(GOA) ، خوارزمية التحسين الجيني(GA) ، تحسين سرب الجسيمات(PSO) ، وتحسين الأسد الاصطناعي(ALO) ، وخوارزمية تحسين الفراشة(BOA) ، وخوارزمية العفن الغروي .(SMA) أكدت معدلات النتائج التي اجريت على 30 تجربة هيمنة EWOA على خوارزميات التحسين الأخرى في جوانب مختلفة، مثل الدقة (0.99987%)، وقيمة اللياقة البدنية (0.00084511%)، واختيار الميزات (متوسط 3.97 ميزة).