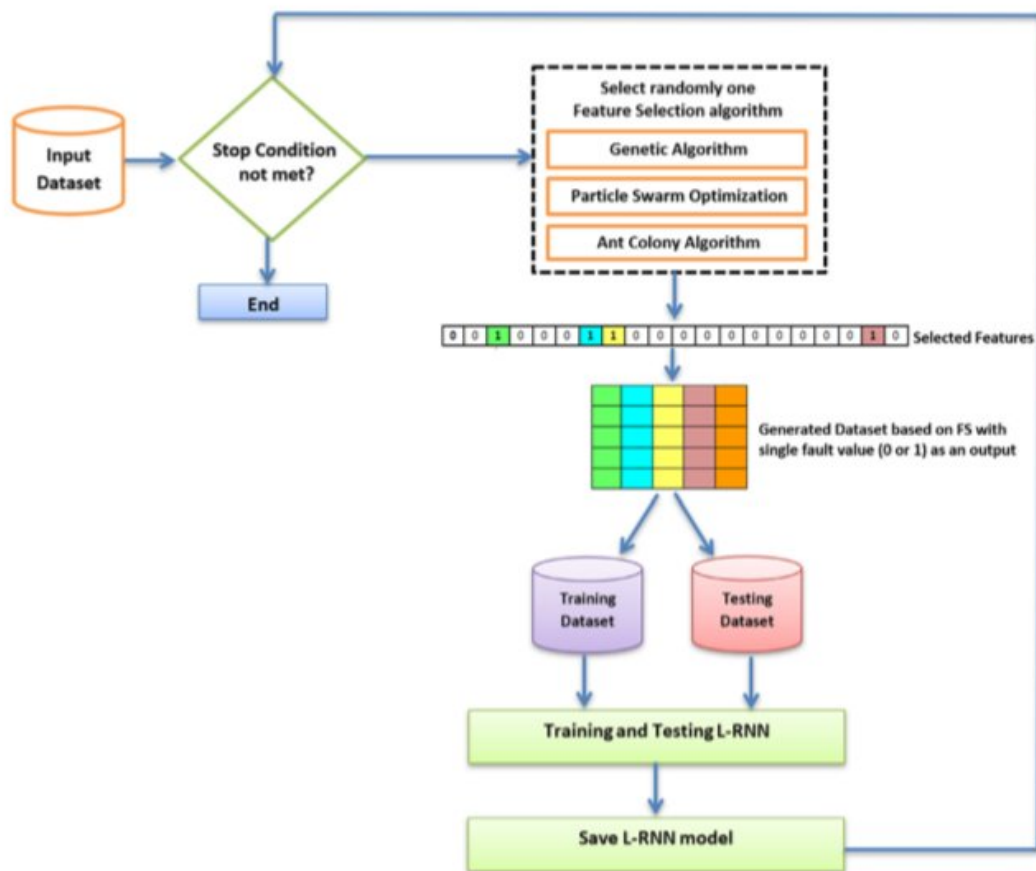


# A new approach for software fault prediction using feature selection

January 10 2019, by Ingrid Fadelli



A pictorial diagram of the proposed methodology. Credit: Turabieh, Mafarja & Li.

Researchers at Taif University, Birzeit University and RMIT University

have developed a new approach for software fault prediction (SFP), which addresses some of the limitations of existing machine learning SFP techniques. Their approach employs feature selection (FS) to enhance the performance of a layered recurrent neural network (L-RNN), which is used as a classification tool for SFP.

Software fault prediction (SFP) is the process of predicting modules that are prone to faults in newly developed [software](#). Predicting faults in [software components](#) before they are delivered to the end user is of key importance, as it can save time, effort and inconvenience associated with identifying and addressing these issues at a later stage.

In recent years, machine learning techniques such as neural networks, logistic regression, support vector machines and ensemble classifiers have proved to be very effective in tackling SFP. However, due to the huge pool of data that can be obtained by mining software historical repositories, it is possible to encounter features that are not related with the faults. This can sometimes mislead the [learning algorithm](#), consequently decreasing its performance.

Feature selection (FS) is a technique that can help to eliminate these unrelated features without impairing the performance of the machine learning [algorithm](#). In machine learning, feature selection entails selecting a subset of relevant features (i.e. predictors) to be used in a particular model. FS can reduce the dimensionality of data; removing irrelevant and redundant data.

In their paper, published in [Expert Systems with Applications](#), the research team at Taif University, Birzeit University and RMIT University proposed a new FS approach to enhance the performance of a layered recurrent neural [network](#) (L-RNN) for SFP. The researchers employed three different wrapper FS algorithms iteratively: binary genetic algorithm (BGA), binary particle swarm optimization (BPSO), and

binary ant colony optimization (BACO).

"We have proposed an iterated feature selection algorithm with a layered recurrent neural network for solving the software faults prediction problem," the researchers wrote in their paper. "The proposed algorithm is able to select the most important software metrics using different feature selection algorithms. The classification process is carried out by a layered recurrent neural network."

The researchers evaluated their approach on 19 real-world software projects from the PROMISE repository and compared their results with those obtained using other state-of-the-art approaches, including Naïve Bayes (NB), artificial [neural networks](#) (ANNs), [logistic regression](#) (LR), the k-nearest neighbors (k-NN) and C4.5 decision trees. Their approach outperformed all other existing methods, achieving an average classification rate of 0.8358 over all datasets.

"The obtained results support our claim of the importance of feature selection in building a high-quality classifier rather than using a fixed set of features or all features," the researchers explained in their paper. "For future work, we plan to investigate the performance of different classifiers such as genetic programming to build a computer model that is able to predict faults based on selected metrics."

**More information:** Hamza Turabieh et al. Iterated feature selection algorithms with layered recurrent neural network for software fault prediction, *Expert Systems with Applications* (2018). [DOI: 10.1016/j.eswa.2018.12.033](#)

© 2019 Science X Network

Citation: A new approach for software fault prediction using feature selection (2019, January 10)

retrieved 10 April 2024 from

<https://techxplore.com/news/2019-01-approach-software-fault-feature.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.