International Journal of Software Science and Computational Intelligence
Volume 18 • Issue 1 • January-December 2026

Adaptive 1-D CNN Using LSelect Feature Selection for Predicting Software Faults

Tamanna Mishra
Guru Jambheshwar University of Science and Technology, India

Sanjay Misra
https://orcid.org/0000-0002-3556-9331
Institute for Energy Technology, Halden, Norway
Received: November 6th, 2025 | **Accepted:** January 9th, 2026

ABSTRACT

Software Fault Prediction (SFP) is the process of predicting fault-prone software constructs during the initial phases of software development. Deep Neural Networks (DNN) have been hugely successful in the field of computer vision, audio, etc., where the input data is correlated spatially and temporally. In contrast, SFP operates on tabular data, rows of software metrics that lack the inherent spatial structure exploited by convolutional architectures in image domains. The authors have tried to remodel the most successful Convolutional Neural Network (CNN) for tabular data. A novel framework is proposed employing a tree-based feature selection technique, LSelect, to find the most significant features and an adaptive 1-dimensional Convolutional Neural Network (ACNN) for the classification task, which selects an optimal learning rate automatically. ACNN converts the tabular data (1-D) into 2-D using adaptive pooling layers, thereby forming an image from 1-D data. The framework classification results (Area under Curve) are compared with nine state-of-the-art algorithms, such as XGBoost, LightGBM, etc., and performance is validated using the Bayesian Signed Rank Test. It is found that the proposed framework performs comparably with the state-of-the-art methods with reduced model complexity. Also, the LSelect feature selection technique improves average model performance by 1.3%.

KEYWORDS

CNN, Tabular Data, Software Fault Prediction, Adaptive Learning Rate, OnecycleLR, Skip Connections, LSelect, Bayesian Statistical Validation, AEEEM

INTRODUCTION

Software Fault Prediction (SFP) is a cornerstone of modern software quality assurance, aiming to enhance software reliability and optimize resource allocation by identifying potentially defective modules early in the development lifecycle. The field has been dominated by traditional Machine Learning (ML) models such as Support Vector Machines (SVM) (Chang & Lin, 2011), Decision Trees (DT) (*Leo Breiman - Classification and Regression Trees-CRC Press (2017)*, n.d.), Naïve Bayes (NB) (ZHANG, 2005), K-Nearest Neighbors (KNN) (Altman, 1992), and more advanced ensemble techniques like XGBoost (Chen & Guestrin, 2016)and LightGBM (LGB) (Ke et al., n.d.).

Despite their widespread use, these models have well-documented limitations that hinder their practical application. A primary challenge lies in the quality of SFP datasets, which are often plagued

DOI: 10.4018/IJSSCI.399476

by severe class imbalance, noise, and irrelevant features, leading to models that are biased and perform poorly in real-world scenarios. Furthermore, many existing models lack generalizability, showing high accuracy on one project's data but failing to adapt to others, which has created a significant gap between academic research and industrial practice.

Concurrently, Deep Neural Networks (DNNs) have achieved state-of-the-art performance across numerous domains, driven by the increasing computational power of Graphics Processing Units (GPUs) (X. Li et al., 2020; Popov et al., 2019; Sercan et al., 2021). Their strength lies in automatically learning hierarchical feature representations from data with inherent spatial, temporal, or semantic structures, as found in images and natural language (Borisov et al., 2021; Shwartz-Ziv & Armon, 2021). However, the application of DNNs, particularly Convolutional Neural Networks (CNNs), to tabular data—the standard format for SFP—has been limited due to the lack of these inherent relationships (Devlin et al., n.d.). Deep Neural Networks (DNNs) have demonstrated immense success on data with strong intrinsic structures, such as the spatial correlation in images or temporal sequences in audio. In contrast, software defect prediction (SFP) operates on tabular data, where each row represents a software module defined by a set of metrics. This data format lacks the explicit spatial or temporal ordering found in images or text. Furthermore, these software metrics are not mutually independent; for instance, object-oriented metrics, such as the Chidamber-Kemerer (CK) suite, are known to exhibit high correlation, as multiple metrics often capture overlapping aspects of software complexity, coupling, and cohesion. This multicollinearity poses a challenge for traditional statistical models, necessitating feature engineering or selection.

This research aims to bridge the gap by developing a novel classification architecture, the "Adaptive CNN" (ACNN), specifically designed to process tabular SFP data. The core of our methodology is a technique that transforms one-dimensional tabular data into a two-dimensional image-like representation, enabling the CNN to leverage its powerful pattern recognition capabilities. To combat the "curse of dimensionality" and improve model efficiency, we also introduce ***LSelect***, a novel feature selection method based on the LightGBM (LGB) algorithm.

To highlight the novelty, practical relevance, and research value of our work, the key contributions are summarized as follows:

- We propose a novel and practical framework for Software Fault Prediction (SFP) that combines Adaptive 1D Convolutional Neural Networks (ACNN) with a LightGBM-based feature selection method (LSelect), addressing key challenges such as high-dimensional tabular data, class imbalance, and model generalization.
- The ACNN transforms 1D tabular data into 2D representations using adaptive pooling, enabling effective feature learning in domains traditionally unsuitable for CNNs. It also integrates automatic learning rate tuning via OneCycleLR, reducing manual configuration and improving convergence.
- The LSelect method filters out noisy and irrelevant metrics, reducing feature dimensionality and improving learning efficiency. Its integration with ACNN leads to a more focused and noise-resilient model, with an average AUC improvement of 1.3%.
- Our framework achieves comparable or superior performance to nine state-of-the-art classifiers (e.g., XGBoost, LightGBM) while maintaining lower model complexity and better cross-dataset generalizability.
- Robustness and statistical significance of the results are validated using the Bayesian Signed Rank Test, ensuring reliability across varied SFP datasets.
- This work presents a novel and automated deep learning pipeline tailored for SFP, offering a scalable and generalizable solution that can assist practitioners in early fault detection with minimal manual intervention.

Novelty and contribution. Recent studies have explored the use of Convolutional Neural Networks for software defect prediction, including 1D-CNNs over metric vectors, multi-channel CNNs, and