SIMATS ENGINEERING
Engineer to Excel
TECH STAR SUMMIT 2024
Name: Mr. Yalaka Nikhil Reddy
Register Number: 192211188
Guided by E.K.Subramanian

Machine Learning based Software Fault Prediction in Software Efficiency using Artificial Neural Network Comparing with Random Forests for Improved Accuracy

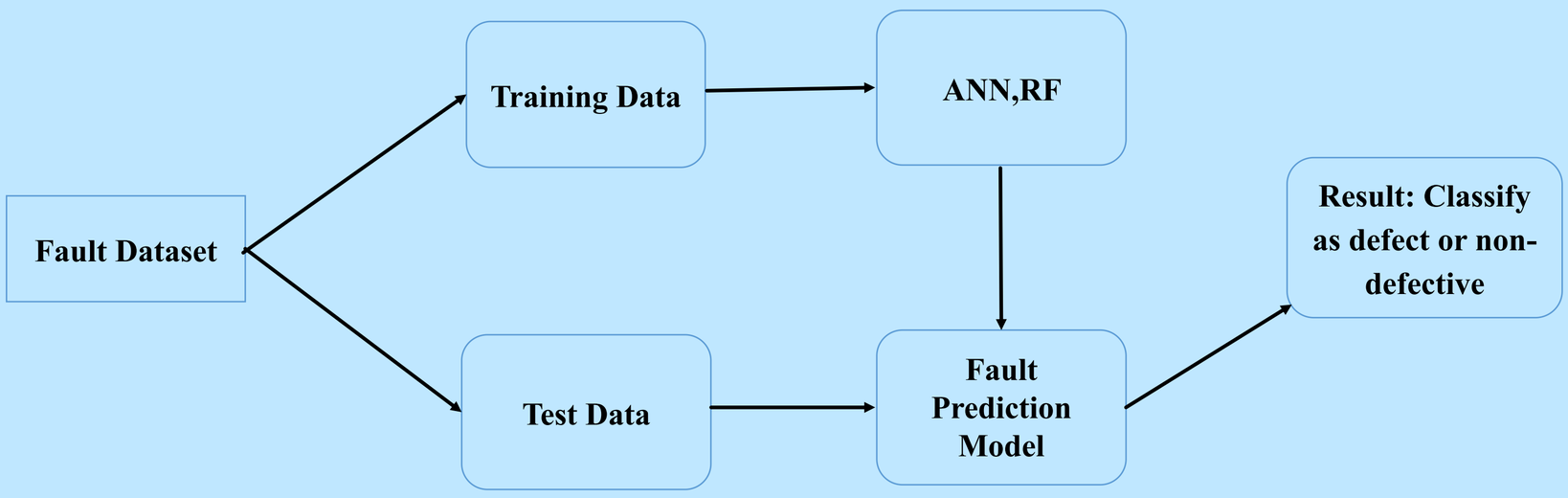
INTRODUCTION

- The aim of this study is to utilize machine learning techniques, specifically Artificial Neural Network (ANN) and Random Forest (RF), for software fault prediction to enhance software efficiency.
- Software fault prediction is crucial in ensuring the reliability and efficiency of software systems. By accurately identifying potential faults, developers can proactively address issues, leading to improved performance, reduced downtime, and enhanced user experience.
- The application of machine learning algorithms such as ANN and RF in software fault prediction offers a data-driven approach to identify patterns and anomalies within the software code, enabling timely interventions to prevent faults from causing system failures.
- A powerful machine learning algorithm inspired by the human brain's neural networks. It can learn complex patterns in data and make predictions with high accuracy.
- This study compares the performance of ANN and RF in terms of accuracy, precision, recall, and F1 score to determine the most effective approach for software fault prediction . Software fault prediction dataset in kaggle includes metrics like code lines, complexity, defect density, enabling robust model training.

Software fault prediction

MATERIALS AND METHODS

- The programming lab of the Saveetha Institute of Medical and Technical Sciences was the site of this investigation. There are two groups. There are two versions of software fault prediction; one is ANN and the other is RF. The sample size was determined using Clinal. Test statistic with a 0.05 threshold, 80% power, and 95% confidence interval the optimal sample size.
- Intel Core i7 processor, 512 GB of RAM, Google Collab, Jupiter Notebooks, and Windows 11th generation. In order to determine accuracy, we conducted statistical analyses using IBM SPSS.



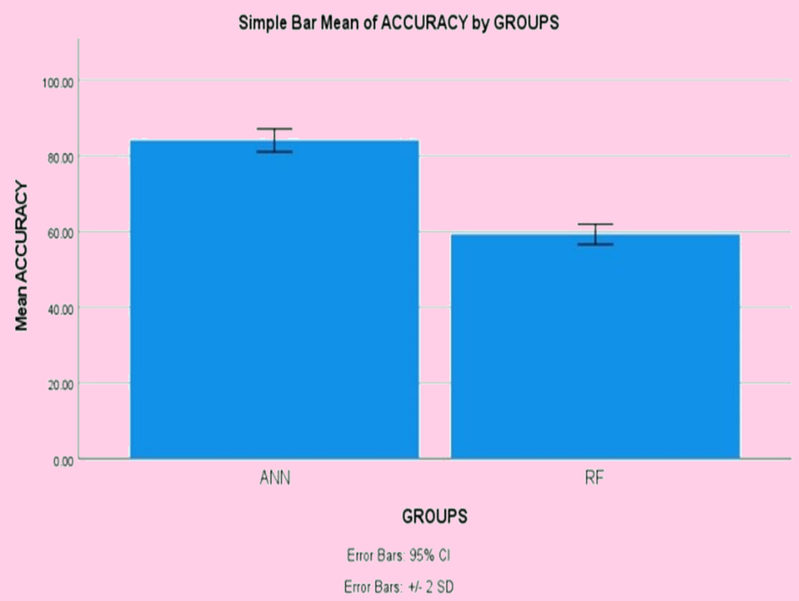
Software Fault Prediction

RESULTS

- Algorithm for software fault prediction using ANN The T-tests for independent samples The following table shows the difference in significance, mean, and standard error between the ANN and Random Forest approaches, as determined by SPSS analysis ($p = 0.001$) ($p<0.05$).
- The results of the ANN and RF algorithms were input into SPSS in order to find the average, dispersion, and error for the two networks. ANN attained 84.10% accuracy, with RF and 59.30% also reaching their goals.
- Using a 95% confidence interval and a standard deviation of +/- 2 as parameters, SPSS calculated an accuracy of 84.10% for the suggested method ANN and 59.30% for the RF.

Statistical Analysis of the difference in significance, mean, and standard error between the Artificial Neural Network and Random Forest approaches

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	ANN	10	84.10	1.52388	0.48189
	RF	10	59.30	1.33749	0.42295



Shows the comparison between Artificial Neural Network and Random Forests with accuracy of 84.10% and 59.30%.

DISCUSSION AND CONCLUSION

- Based on T-test statistical analysis with a significance value of $p=0.001$ (independent sample T-test $p<0.05$), it is observed that there is a statistically significant difference between the performance of Artificial Neural Network (ANN) and Random Forest (RF) algorithms.
- The overall accuracy of the Random Forest Algorithm is 84.10%, which outperforms other algorithms, including Support Vector Machine (SVM) with an accuracy of 59.30%.
- Further research could explore fine-tuning these algorithms for specific software development environments or types of software projects. Additionally, investigating ensemble methods or hybrid approaches could enhance prediction accuracy further.
- Factors such as the quality and quantity of data, feature selection, model hyperparameters, and preprocessing techniques can influence the accuracy of software fault prediction models.
- One limitation of this study may be the size and diversity of the dataset used. Additionally, the performance of machine learning algorithms can be influenced by the choice of features and the preprocessing steps applied to the data. This study contributes to the understanding of machine learning-based software fault prediction, highlighting the importance of choosing appropriate algorithms for maximizing accuracy.

BIBLIOGRAPHY

- "Machine Learning Based Methods for Software Fault Prediction: A Survey." 2021. Expert Systems with Applications 172 (June): 114595.https://doi.org/10.1016/j.eswa.2021.114595
- Matloob, Faseeha, Taher M. Ghazal, Nasser Taleb, Shabib Aftab, Munir Ahmad, Muhammad Adnan Khan, Sagheer Abbas, and Tariq Rahim Soomro. n.d. "Software Defect Prediction Using Ensemble Learning: A Systematic Literature Review." Accessed March 18, 2024.http://dx.doi.org/10.1109/ICBATS54253.2022.9759059
- Assim, Marwa, Qasem Obeidat, and Mustafa Hammad. n.d. "Software Defects Prediction Using Machine Learning Algorithms." Accessed March 18, 2024.http://dx.doi.org/10.1109/ICDABI51230.2020.9325677
- Sun, Zhongbin, Qinbao Song, and Xiaoyan Zhu. n.d. "Using Coding-Based Ensemble Learning to Improve Software Defect Prediction." Accessed March 18, 2024. https://doi.org/10.1145/3640331