SIMATS ENGINEERING

Engineer to Excel

SAVEETHA TRANSDISCIPLINARY ANNUAL RESEARCH SUMMIT · STAR SUMMIT 2024 · SIMATS ENGINEERING

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 Logistic Regression for Improved Accuracy

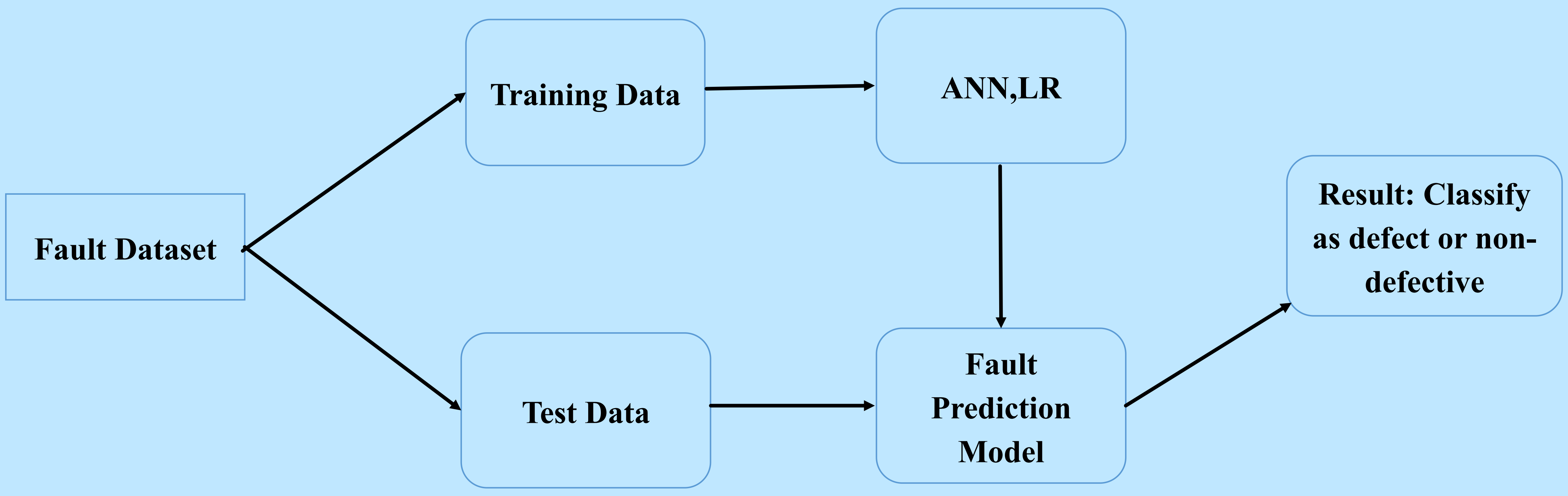
INTRODUCTION

- The aim of this study is to predict software faults in software efficiency using machine learning techniques, specifically comparing the performance of Artificial Neural Network (ANN) with Logistic Regression.
- Software faults can lead to significant issues such as system crashes, security vulnerabilities, and performance degradation. Predicting these faults beforehand can help in proactively addressing them, reducing downtime, improving user experience, and minimizing maintenance costs.
- The application of machine learning for software fault prediction can benefit various industries relying on software systems, including technology, finance, healthcare, and manufacturing.
- Both Artificial Neural Network (ANN) and Logistic Regression are popular machine learning algorithms for classification tasks. ANN, inspired by the structure of the human brain, can capture complex relationships in data, while Logistic Regression is simpler and interpretable, making it easier to understand the influence of input features on the output.
- Software fault prediction dataset in kaggle includes metrics like code lines, complexity, defect density, enabling robust model training. These features serve as input variables for training the machine learning models, while the output variable is the presence or absence of software faults.

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 LR. 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.



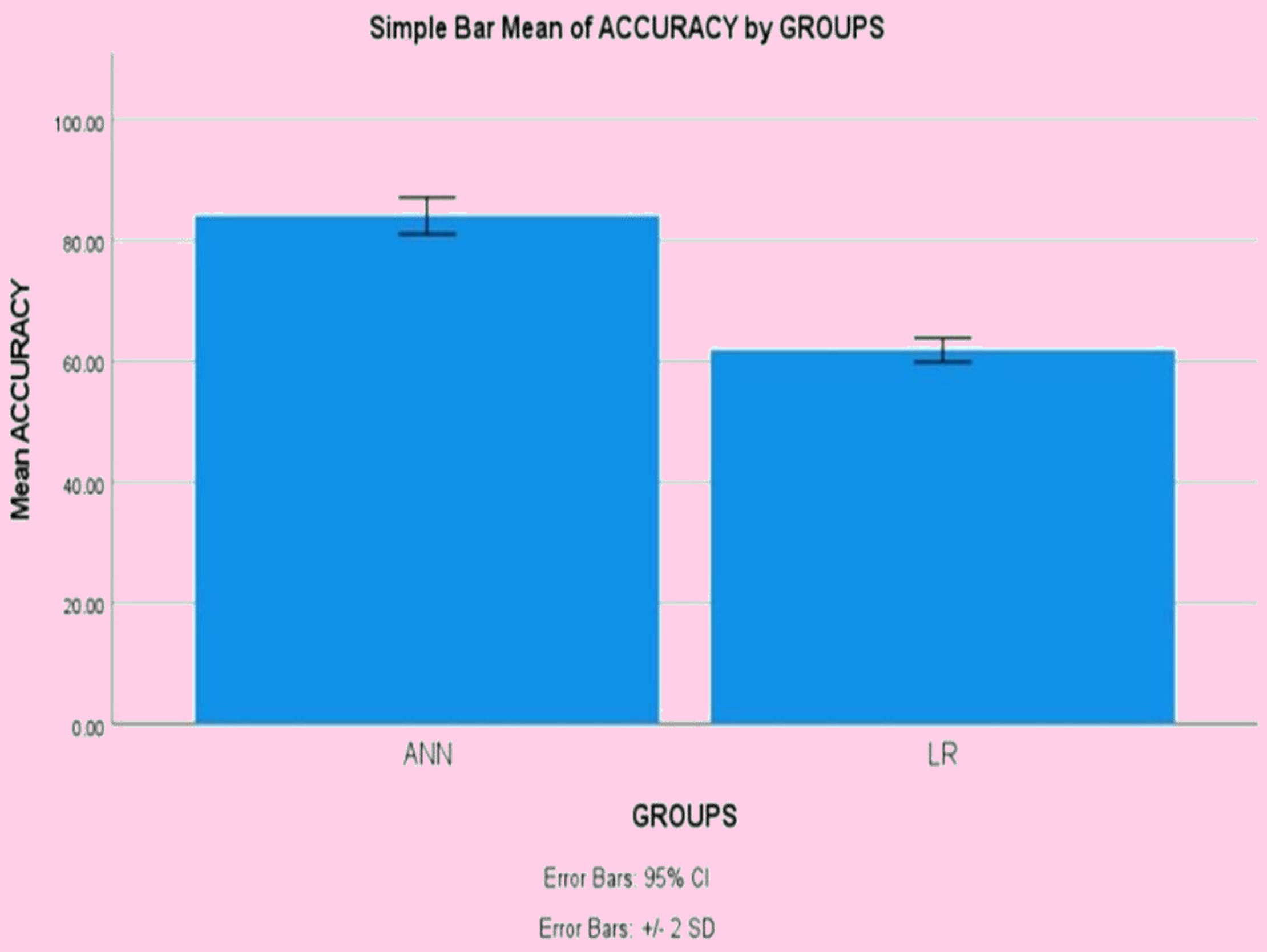
Software Fault Prediction

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 LR approaches, as determined by SPSS analysis (p = 0.001) (p<0.05).
- The results of the ANN and LR algorithms were input into SPSS in order to find the average, dispersion, and error for the two networks. ANN attained 84.10% accuracy, with LR and 61.90% 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 61.90% for the LR .

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

	Algorithm	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	ANN	10	84.10	1.52388	0.48189
	LR	10	61.90	0.99443	0.31447



Shows the comparison between the Artificial Neural Network and Logistic Regression with accuracy of 84.10% and 61.90%.

DISCUSSION AND CONCLUSION

- Based on T-test statistical analysis with a significance value of p=0.001 (independent sample T-test p<0.05), there is a statistically significant difference between the group 1 (Artificial Neural Network) and group 2 (Logistic Regression). The overall accuracy of the models is as follows:
- Artificial Neural Network : 84.10%
- Logistic Regression: 61.90%
- This indicates that the ANN model achieves a higher accuracy compared to Logistic Regression in predicting software faults in software efficiency.
- Moving forward, there is a scope for further enhancing the predictive accuracy of software fault prediction models. This could involve exploring advanced machine learning techniques, such as ensemble methods like Random Forest, and fine-tuning the models for specific software development environments or domains.
- Several factors can influence the accuracy of software fault prediction models, including the quality and quantity of training data, feature selection, model architecture, and hyperparameter optimization. Future research should consider these factors comprehensively to improve prediction performance.
- Additionally, the performance of the models may be influenced by factors not accounted for in this study.

BIBLIOGRAPHY

- Li, Jian, Pinjia He, Jieming Zhu, and Michael R. Lyu. n.d. "Software Defect Prediction via Convolutional Neural Network." Accessed March 18, 2024.https://doi.org/10.1145/3460319.3464819
- "An Efficient Convergence-Boosted Salp Swarm Optimizer-Based Artificial Neural Network for the Development of Software Fault Prediction Models." 2023. Computers & Electrical Engineering 111 (October): 108923.http://dx.doi.org/10.1016/j.compeleceng.2023.108923.
- Azam, Muhammad, Muhammad Nouman, and Ahsan Rehman Gill. 2022. "Comparative Analysis of Machine Learning Technique to Improve Software Defect Prediction." KIET Journal of Computing and Information Sciences 5 (2). https://doi.org/10.51153/kjcis.v5i2.96.https://doi.org/10.51153/kjcis.v5i2.96.
- Babu, Rathish Babu Thirukonda Krishnamoorthy, Suresh Sivasubramanian, and Sankarram Natarajan. 2022. "MLPNN-RF: Software Fault Prediction Based on Robust Weight Based Optimization and Jacobian Adaptive Neural Network." Concurrency and Computation: Practice & Experience 34 (21): e7122. https://doi.org/10.1002/cpe.7122.