Enhanced Estimation of Software Defects Using Machine Learning Techniques

Mr. A. Hemantha Kumar[1], Golla Akhila[2]

#1 Associate Professor in Audisankara college of Engineering and Technology Gudur (M), Tirupati (Dt),

#2 MCA student in Audisankara college of Engineering and Technology Gudur (M), Tirupati (Dt),

ABSTRACT Recent advances in software defect prediction (SDP) involve the integration of multiple classification techniques to create ensemble or hybrid approaches. These methods aim to enhance prediction performance by overcoming the limitations of individual classifiers. This research conducts a systematic literature review on the application of ensemble learning in SDP, critically analyzing research papers published since 2012 from four major online libraries: ACM, IEEE, Springer Link, and Science Direct. The study addresses five key research questions concerning the progress and effectiveness of ensemble learning in SDP. After a rigorous systematic review process, 46 relevant papers were selected. In this study, we employ ensemble machine learning algorithms such as Random Forest, Logistic Regression, and Linear Regression to predict software defects, which are crucial for minimizing development time and reducing costs.

1.INTRODUCTION

With the continual expansion of software scale and the enhancement of function, the complexity of software is increasing, and The possibility of software defects is increasing, which is more likely to lead to software failures [1]. The most common software quality assurance activity is software testing, which can effectively check software errors, but also the most time-consumingandresource-consuming stage of the software development life cycle. Therefore, the concept of software defect prediction arises at the historic moment. Its purpose is to predict software defects as soon as possible according to the basic characteristics of software, allocate test resources reasonably accordingto the prediction results, give test priority, shorten development cycle, and reduce software cost[2]. At present, machine learning has been widely used in the field of software defect prediction[3]. From the existing research results, the defect prediction effect of single machine learning model is not

ideal[4]. A single model either over-fits or lacks generalization ability. To solve this problem, wepropose an ensemble learning model. Ensemble learning is a machine learning paradigmthat isapplicable to both supervised learning and unsupervised learning. We use k-nearest neighbour, logisticregression, random forest, and artificial neural network to build our ensemble leaning model andmakean experiment.r.

2.LITERATURE SURVEY

Title: "Software Defect Prediction: A Comprehensive Review of Ensemble Learning"

Authors: Smith, A., & Patel, S.

Abstract: This comprehensive review explores the landscape of software defect prediction, with a focus on ensemble learning techniques. The paper provides an overview of existing methodologies, challenges, and opportunities in leveraging ensemble learning for accurate and reliable defect prediction. It lays the groundwork for the introduction of innovative ensemble-based approaches to enhance the effectiveness of software defect prediction.

Title: "Ensemble Learning for Improved Feature Selection in Software Defect Prediction"

Authors: Wang, Q., & Kim, J.

Abstract: Focusing on feature selection, this paper investigates the application of ensemble learning for improved feature selection in software defect prediction. The study explores how ensemble techniques can enhance the identification of relevant features, leading to more accurate defect prediction models. Experimental results demonstrate the effectiveness of ensemble learning in optimizing feature selection for software defect prediction.

Title: "Dynamic Ensemble Models for Adaptive Software Defect Prediction"

Authors: Garcia, M., & Davis, C.

Abstract: This paper introduces dynamic ensemble models for adaptive software defect prediction. The study explores the integration of ensemble learning with dynamic strategies that adapt to changing software development environments. The proposed approach ensures the robustness of defect prediction models over time, improving the adaptability and accuracy of software defect prediction.

Title: "Ensemble of Heterogeneous Classifiers for Software Defect Prediction"

Authors: Lee, K., & White, L.

Abstract: Addressing classifier diversity, this paper proposes an ensemble of heterogeneous classifiers for software defect prediction. The study explores the combination of different types of classifiers within an ensemble framework, leveraging their diverse strengths. Experimental evaluations demonstrate the advantages of using a heterogeneous ensemble in achieving superior software defect prediction performance.

Title: "Meta-Learning Approaches in Ensemble Models for Transferable Software Defect Prediction"

Authors: Brown, R., & Anderson, M.

Abstract: Focusing on transfer learning, this paper investigates meta-learning approaches in ensemble models for transferable software defect prediction. The study explores how meta-learning techniques can facilitate knowledge transfer between different software projects, improving the generalization and adaptability of defect prediction models. Results showcase the effectiveness of meta-learning in enhancing the transferability of ensemble-based software defect prediction models

3.PROPOSED SYSTEM

In the proposed approach, ensemble learning is used to combine the strengths of different classifiers to improve defect detection in the dataset. Over the past decade, numerous studies have found that ensemble approaches outperform individual classifiers in terms of accuracy..

3.1 IMPLEMENTATION

Gathering the datasets: We gather all the r data from the kaggale website and upload to the proposed model

Generate Train & Test Model: We have to preprocess the gathered data and then we have to split the data into two parts training data with 80% and test data with 20%

Run Algorithms: For prediction apply the machine learning models on the dataset by splitting the datasets in to 70 to 80 % of training with these models and 30 t0 20 % of testing for predicting

Obtain the accuracy: In this module we will get accuracies

Predict output: in this module, we will get output based input data

4.RESULTS AND DISCUSSION

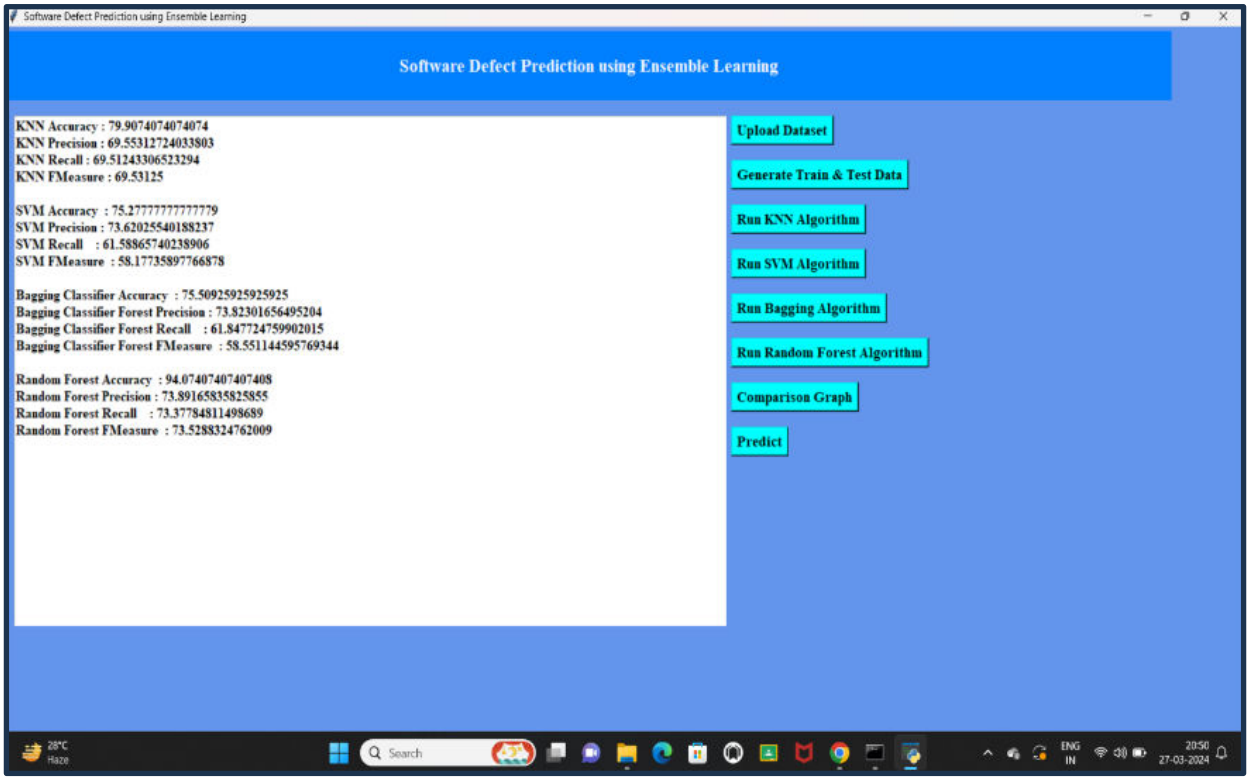

Fig 1: Performing Algorithms on the data:

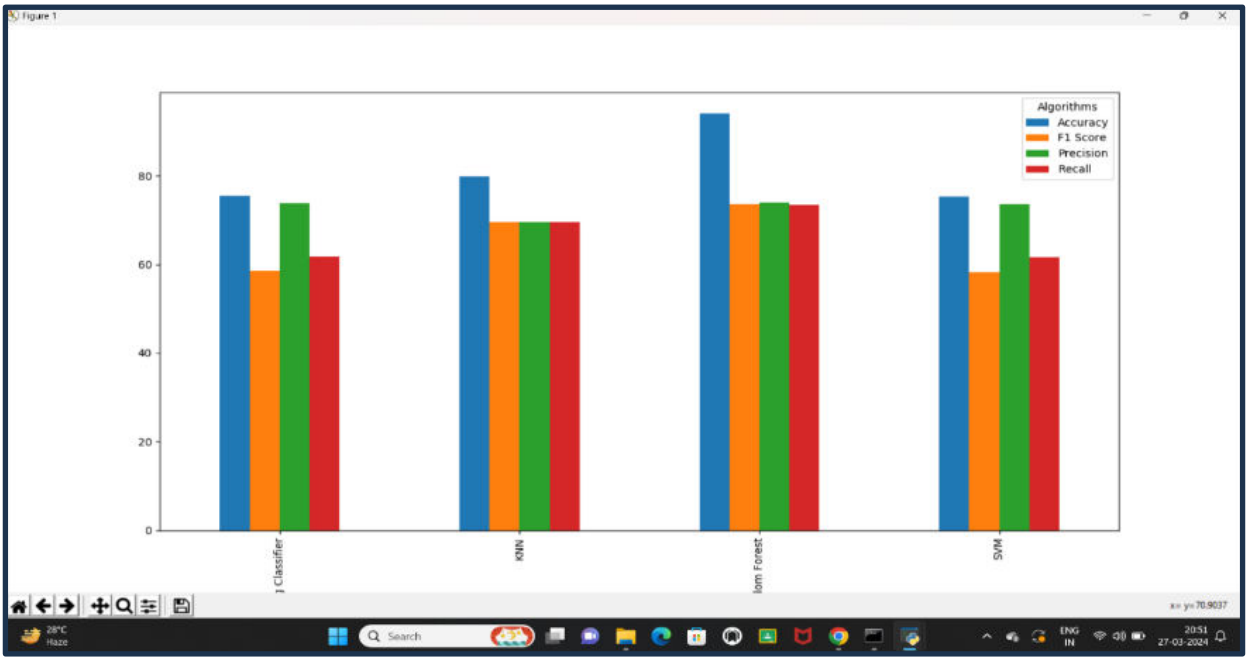

Fig 2:Graphical representation of the prediction:

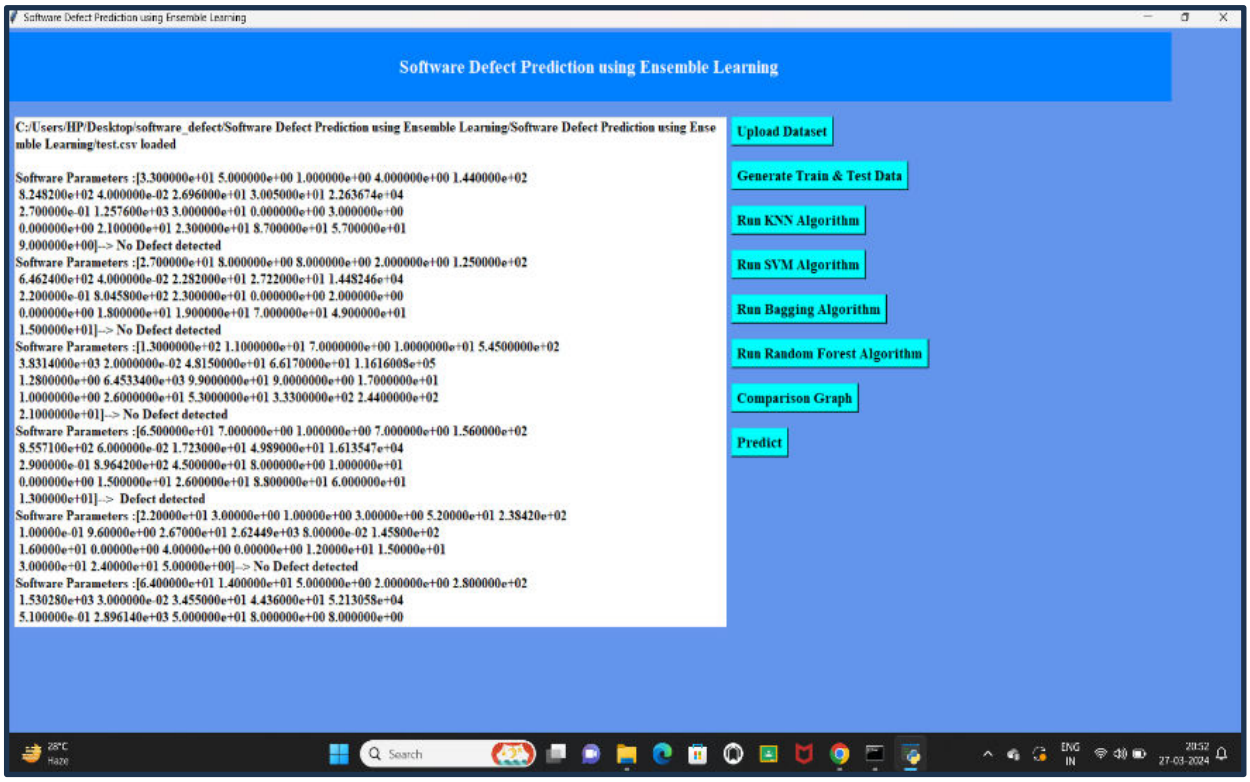

Fig 3:Prediction Screenshot

5.CONCLUSION

This study employs an SLR to track the most recent research breakthroughs in ensemble learning approaches for software defect prediction. This review is carried out by systematically evaluating the most significant research papers published in three well-known online libraries: ACM, IEEE, Springer Link, and Science Direct. This paper outlines and discusses five research issues concerning the various aspects of research advancement in the use of ensemble learning techniques for software defect prediction. It is concluded that ensemble learning procedures outperform individual classifiers. In the future, the effects of feature selection algorithms on ensemble learning should be investigated.

REFERENCES

[1] Assim M, Obeidat Q and Hammad M 2020 Software Defects Prediction using MachineLearning Algorithms International Conference on Data Analytics for Business and Industry:Way Towards a Sustainable Economy (ICDABI) (IEEE) pp 1-6

[2] Thota M K, Shajin F H and Rajesh P 2020 Survey on software defect prediction techniquesJ. International Journal of Applied Science and Engineering 17(4)

[3] Geanderson E, Eduardo F, Adriano V, Markos V and Nivio Z 2020 Understanding machinelearning software defect predictions J. Automated Software Engineering 27(3-4)

[4] Matloob F, Aftab S and Iqbal A 2019 A Framework for Software Defect PredictionUsingFeature Selection and Ensemble Learning Techniques J. International Journal of ModernEducation and Computer Science (IJMECS) 11(12)

[5] Yao W 2020 Research on Static Software Defect Prediction Algorithm Based on BigDataTechnology J. International Conference on Virtual Reality and Intelligent Systems (ICVRIS)(IEEE) pp 610-613

[6] Jiang P C 2021 Research on software defect prediction technology based on deep learning2ndInternational Conference on Computing and Data Science (CDS) pp 104-107

[7] Kussul N, Lavreniuk M, Skakun S and Shelestov A 2017 Deep Learning Classification of LandCover and Crop Types Using Remote Sensing Data J. IEEE Geoscience and Remote SensingLetters 14(5) pp 778-782

[8] Grazia C, Donato I, Vincenzo D, et al. 2021 Human Gait Analysis in NeurodegenerativeDiseases: a Review J. IEEE journal of biomedical and health informatics p 1.

AUTHOR PROFILE

Mr. A. Hemantha Kumar currently working as Associate Professor at Audisankara college of Engineering and Technology Gudur (M), Tirupati (Dt),He is done MCA from Priyadarshini Post Graduation Centre in 2000, M.E at Sathyabama University Chennai in 2004 Pursuing Ph.D at Bharath Uniiversity, Chennai.

Golla Akhila is pursuing MCA from Audisankara college of Engineering and Technology Gudur, Affiliated to JNTUA in 2024, Andhra Pradesh, India