ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

# Software Defect Prediction Using An Intelligent Enssemble Based Model

**Mr C.Krupa Sagar Reddy[1], Akkem Venkata Naga Padmini[2], DasariAlekhya[3], Bachala Nithya Sree[4], Kalle Lakshmi Prasanna[5], Gajjala Vijaya Lakshmi[6]**

[1,2,3,4,5,6]Chaitanya Bharathi Institute of Technology, Proddatur, A.P, India.

krupasagar.challa@gmail.com[1],padduakkem@gmail.com[2],dasarialekya123@gmail.com[3],nithyasreebachala @gmail.com[4],prasannakalle.1803@gmail.com[5], vijayalakshmigajjala197@gmail.com[6]

## ABSTRACT

In the software engineering field, defect prediction is a highly active area of research. Bridging the gap between software engineering and data mining is crucial for the success of software development. Predicting software defects involves identifying errors in source code before the testing phase. Various methods are employed for defect prediction, including clustering, statistical methods, hybrid algorithms, neural network-based metrics, black box testing, white box testing, and machine learning techniques. This research introduces the use of feature selection to enhance the accuracy of machine learning classifiers in predicting software defects for the first time. The study aims to improve prediction accuracy using five publicly available NASA datasets: CM1, JM1, KC2, KC1, and PC1. Feature selection is applied in conjunction with machine learning techniques such as Random Forest, Logistic Regression, Multilayer Perceptron, Bayesian Network, Rule ZeroR, J48, Lazy IBK, Support Vector Machine, Neural Networks, and Decision Stump, demonstrating higher defect prediction accuracy compared to methods without feature selection. The WEKA (Waikato Environment for Knowledge Analysis) tool is utilized for data refinement, preprocessing, and classifier application. Additionally, statistical analyses are performed using the Minitab statistical tool. The study's findings indicate that defect prediction accuracy is significantly improved with feature selection compared to without feature selection

**Keywords:** Software quality, Predictive models, Support vector machines, Prediction algorithms, Random forests.

## I. INTRODUCTION

Software defects refer to unexpected performance issues in response to a client's needs. These defects, often identified by software testers, arise due to errors, faults, or failures in the software development process. Predicting software defects is a crucial aspect of software engineering, as it helps detect malfunctioning modules early in the development cycle. Traditional defect detection methods such as code reviews, beta testing, and system testing become challenging as software complexity grows. Hence, defect prediction models play a vital role in identifying defects early, improving software quality, and reducing maintenance costs.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

Software defect prediction has gained popularity due to its direct impact on software quality. Defective modules can lead to cost overruns, project delays, and increased maintenance expenses. Two fundamental methods of software quality assurance are defect detection and defect prevention. While defect detection addresses existing flaws, defect prevention focuses on minimizing potential faults early in development. Predicting defects before deployment enables better resource allocation, leading to cost savings and high-quality software. By using prediction models, developers can proactively enhance software quality, reducing defects and ensuring user satisfaction.

Machine learning has become a powerful tool for software defect prediction. By analyzing past data, machine learning algorithms can recognize patterns, predict defects, and improve decision-making with minimal human intervention. These algorithms enhance software quality by identifying potential issues early, allowing organizations to allocate testing efforts more effectively. Classification, clustering, and regression techniques are commonly used in defect prediction models. Machine learning-based prediction also reduces rework, ensuring better efficiency in software development.

The application of machine learning in defect prediction helps organizations prioritize testing, optimize resources, and enhance software reliability. By identifying high-risk areas, developers can address issues before they affect end-users, improving customer satisfaction and reducing maintenance efforts. This research contributes to improving the accuracy of machine learning algorithms in defect prediction, leading to better software quality assurance.

## II. LITERATURE SURVEY

### 1. Early Machine Learning Approaches :

Langley and Carbonell (1984) introduced different machine learning methods. Dietterich (2009) explored machine learning applications in ecosystem informatics and sustainability.

### 2. Classification Algorithms for Medical Diagnosis :

Ramana, Babu,and Venkateswarlu (2011) evaluated classification algorithms for liver disease diagnosis.

## III. EXISTING SYSTEM

Defect prediction in software engineering has been extensively studied using machine learning (ML) and software metrics. Early research by Benton and Neil (1999) introduced size and complexity metrics, estimating that each KLOC contains about 23 defects. Various ML approaches, such as neural networks, support vector machines (SVM), and ensemble methods, have been explored to enhance defect prediction accuracy. Vanmali et al. demonstrated that neural networks outperformed other models using the PROMISE dataset, while Askari and Bardsiri combined evolutionary techniques with SVM for improved precision. Additionally, Gray et al. emphasized the importance of data quality, highlighting that noisy or missing values could significantly impact prediction results.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

Several studies investigated cost-effective strategies for defect detection. Biçer et al. focused on static code measures, showing that flaw identifiers consistently produce reliable results across applications. Prasad et al. explored supervised, unsupervised, and semi-supervised ML techniques, with decision trees, Bayesian networks, and clustering proving effective in predicting defects. Kumar and Shukla proposed a fuzzy logic-based model that utilizes multiple metrics for early defect detection, while Chandra Yadav et al. applied classification and association rule mining to enhance error identification with minimal testing resources.

Researchers have also compared various ML models for defect prediction across datasets. Ramana et al. tested classification techniques on liver patient datasets, revealing that K-Nearest Neighbor , backpropagation, and SVM yielded superior results. Hammouri et al. examined debugging datasets using Naive Bayes, artificial neural networks, and decision trees, concluding that ML models outperform traditional approaches. Lastly, Memon et al. reviewed defect prediction and prevention mechanisms, discussing pattern-based and graph mining techniques to minimize software failures and improve software quality.

**Advantages of the Existing System:**

1. **Improved Prediction Accuracy** – The use of ensemble methods (Random Forest, Bayesian Networks) enhances classification accuracy by combining multiple models.
2. **Versatile Machine Learning Approaches** – The system incorporates supervised, unsupervised, and semi-supervised techniques, making it adaptable to different types of data.
3. **Automation of Defect Detection** – Machine learning algorithms reduce manual effort and human error, improving efficiency in defect prediction.
4. **Scalability and Generalization** – Evolutionary approaches combined with SVM learning allow the system to handle large datasets effectively and generalize well to new data.

**Disadvantages of the Existing System:**

1. **High Computational Complexity** – Using multiple machine learning techniques, especially ensemble methods and evolutionary approaches, can require significant computational resources.
2. **Data Dependency** – The system's performance heavily depends on the quality and quantity of training data, making it less effective if data is insufficient or imbalanced.
3. **Difficult Model Selection and Tuning** – Choosing the right combination of supervised, unsupervised, and semi-supervised methods can be complex and require extensive tuning.
4. **Limited Interpretability** – Some machine learning models, such as neural networks and ensemble methods, act as "black boxes," making it difficult to understand how predictions are made.

## IV. PROPOSED SYSTEM

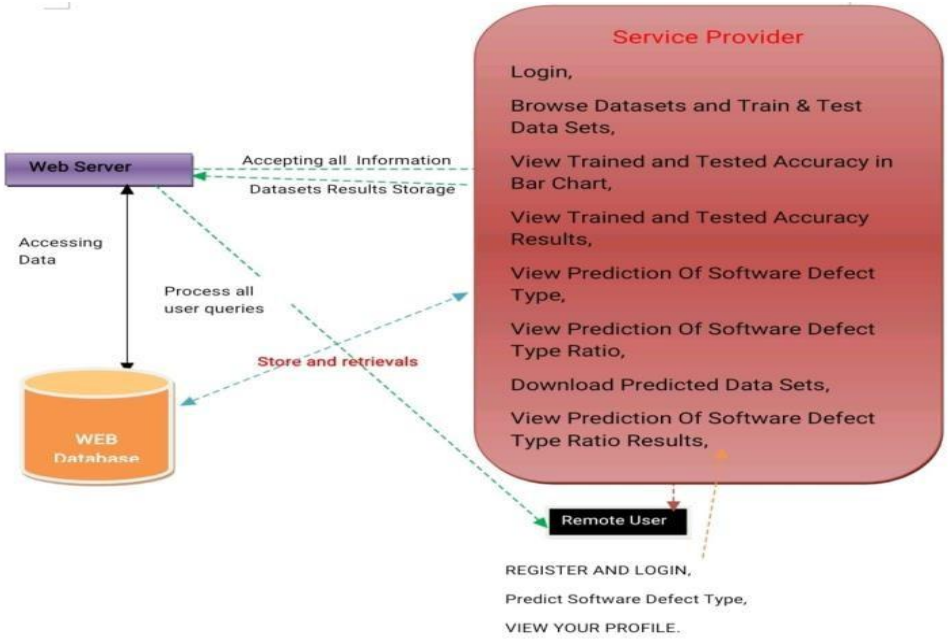
The main contribution of this research is the use of feature selection for the first time to increase the accuracy of machine learning classifiers in defects prediction. The objective of this study is to improve the defects prediction accuracy in five data sets. The machine-learning techniques used in this research are; Random Forest, Logistic Regression, Multi-layer Perceptron, Bayesian Net, Rule ZeroR, J48, Lazy IBK, Support Vector Machine, Neural Networks, and Decision Stump to achieve high defect prediction accuracy as compared to WOFS.

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

**Advantages of the Proposed System:**

1. **Improved Prediction Accuracy** – Feature selection helps eliminate irrelevant data, allowing machine learning models to focus on the most important attributes, leading to higher defect prediction accuracy.
2. **Reduced Computational Complexity** – By minimizing the number of input features, models require less processing power and memory, making them more efficient and faster.
3. **Better Generalization** – Removing redundant features reduces overfitting, ensuring that the model performs well on new, unseen datasets.
4. **Enhanced Interpretability** – A smaller, more relevant feature set makes it easier to analyze and understand which factors contribute most to software defects, aiding in decision-making.

**ARCHITECTURE :**




**MODULES:**

1. **Service Provider** :In this module, the Service Provider has to login by using valid user name and password. After login successful he can do some operations such as Browse Datasets and Train & Test Data Sets, View Trained and Tested Accuracy in Bar Chart, View Trained and Tested Accuracy Results, View Prediction Of Software Defect Type, View Prediction Of Software Defect Type Ratio, Download Predicted Data Sets, View Prediction Of Software Defect Type Ratio Results, View All Remote Users.
2. **View and Authorize Users :**In this module, the admin can view the list of users who all registered. In this, the admin can view the user's details such as, user name, email, address and admin authorize the users.
3. **Remote User** :In this module, there are n numbers of users are present. User should register before doing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user

ISSN: 2582 - 6379
Orange Publications
International Journal for Interdisciplinary Sciences and Engineering Applications
IJISEA - An International Peer- Reviewed Journal
2025, Volume 6 Issue 1
www.ijisea.org

will do some operations like REGISTER AND LOGIN, Predict Software Defect Type, VIEW YOUR PROFILE.

## V. CONCLUSION

Software defect prediction is essential for detecting flaws in source code before testing, especially as software systems become more complex. Traditional testing methods struggle with large-scale projects, making machine learning techniques like classification, clustering, and neural networks valuable for improving defect detection. However, no single approach works universally across datasets. This study enhances defect prediction accuracy by applying feature selection to five NASA datasets (JM1, CM1, KC1, KC2, and PC1) using the WEKA machine-learning workbench. Feature selection improves the Bayesian. Net algorithm's accuracy by 8%, while Logistic Regression achieves over 93% accuracy. Future research can focus on optimizing prediction models by exploring metaheuristic feature selection techniques to identify the most relevant attributes. Addressing data imbalance remains a challenge, as it impacts model performance. Additionally, investigating the effectiveness of deep learning algorithms and ensemble classifiers with different resampling strategies may further enhance accuracy and reliability in software defect prediction.

## REFERENCES:

[1] M. A. Memon, M.-U.-R. Magsi, M. Memon, and S. Hyder, "Defects prediction and prevention approaches for quality software developments. J. Adv. Compute. Sci. Appl., vol. 9, no. 8, pp. 451–457, 2018.
[2] M. Gayathri and A. Sudha, "Software defect prediction system using multilayer perceptron neural network with data mining," Int. J. Recent Technol. Eng., vol. 3, no. 2, pp. 2277–3878, 2014.
[3] R. Malhotra, L. Bahl, S. Sehgal, and P. Priya, "Empirical comparison of machine learning algorithms for bug prediction in open-source software," in Proc. Int. Conf. Big Data Anal. Comput. Intell. (ICBDAC), Andhra Pradesh, India, 2017, pp. 40–45, doi: 10.1109/ICBDACI.2017.8070806.
[4] M. S. Rawat and S. K. Dubey, "Software defect prediction models for quality improvement: A literature study," Int. J. Compute. Sci. Issues, vol. 9, pp. 288–296, Jan. 2012.
[5] I. Singh, "A survey? Data mining techniques in software engineering," Int. J. Res. IT, Manage. Eng., vol. 6, no. 3, pp. 30–34, Mar. 2016.
[6] N. Kalaivani and R. Beena, "Overview of software defect prediction using machine learning algorithms," Int. J. Pure Appl. Math., vol. 118, pp. 3863–3873, Feb. 2018.
[7] M. Dhiauddin and S. Ibrahim, "A prediction model for system testing defects using regression analysis," Int. J. Soft Compute. Softw. Eng., vol. 2, no. 7, pp. 55–68, Jul. 2012.
[8] R. Malhotra and A. Jain, "Fault prediction using statistical and machine learning methods for improving software quality," J. Inf. Process. Syst.,vol. 8, no. 2, pp. 241–262, Jun. 2012.
[9] Hammouri, M. Hammad, M. Alnabhan, and F. Alsarayrah, "Software bug prediction using machine learning approach," Int. J. Adv. Compute. Sci.Appl., vol. 9, no. 2, pp. 78–83, 2018.
[10] M. M. A. Abdallah and M. M. Alrifaee, "Towards a new framework of program quality measurement based on programming language standards. J. Eng. Technol., vol. 7, pp. 1–3, Oct. 2018.