International Journal of Research and Development in Engineering Sciences
ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , Conference Proceedings
International Conference on Emerging Trends in Engineering, Technology and Management 2025
11-12th April 2025 , www.icetm.in , ICETM 2025 , Hybrid Mode
Viswam Engineering College (Autonomous) , Madanapalle , Andhra Pradesh , India
Conference Proceedings , https://doi.org/10.63328/IJRDES-V7C1-ICETM2025
Regular Issue, Vol. 7, ICETM 2025 , Pages: 14 – 18 , ISBN : 978-81-984440-5-9
Received: 15 January 2025 ; *Accepted:* 08 April 2025 ; *Published:* 12 June 2025
https://doi.org/10.63328/IJRDES-V7CIP3
Conference Paper : Original – Action Research – Experimental Research Paper

Towards Generalizable Models in Software Failure Prediction: A Machine Learning Approach

Subbalakshmamma Thaadi [1], **M Anjan Kumar** [2], **D Ramana Reddy** [3]

[1-3] Viswam Engineering College, Madanapalle, Andhra Pradesh, India – 517102

** Corresponding Author : Subbalakshmi Thaadi ; ramaa4u.iitm@gmail.com*

Abstract: The inability to communicate effectively because of communication barriers severely restricts deaf and mute This paper highlights the challenges of generalizing models and dealing with contradictory evidence. Additionally, it explores the potential to enhance software failure prediction by integrating multiple research efforts. Traditional methods of failure prediction could not be highly task-specific due to the fact that not all tasks had access to the same data. To overcome these challenges and achieve better prediction accuracy, you can employ feature selection techniques, data resampling strategies, and machine learning processes. The project includes exploring the use of different datasets and enhancing model training to accelerate problem detection and ensure that solutions are compatible with different software configurations. Software quality assurance methods can be improved and made more adaptable as a direct result of the findings.

Keywords: Software Fault Prediction, Cross-Project Analysis, Imbalanced Data, ML, SQA.

1. Introduction

Software fault prediction (SFP) is a necessary part of software quality assurance because it helps to identify the problematic modules at their early development phase. The reason why traditional approaches to the defect estimation cannot be used in this situation is that they need homogenous data used in a single project. The answer to this problem may be the cross-project software fault prediction (CP-SFP) that minimizes the use of project-specific data and enhances the flexibility of the model incorporating the data of multiple projects. An issue with CP-SFP and a possible reason behind wrong counting is data imbalance that is such that there are few bad modules as compared to good ones.

To become more efficient at forecasting software failures on projects with the help of analytics, you must resolve the issue of data anomalies. In response to the predictive question of non-faulty units, machine learning systems actually perform poorly, as the data of faults are not sufficiently diverse.

Cost-sensitive learning, hybrid techniques and resampling are some of the solutions to this problem. It is possible to use feature selection and transfer learning methods to produce interdependent projects and enhance the overall model generalizability and give the correct predictions in other software environments. Projects, the processes of the development and coding standards are all prone to change and so the accuracy of the predictions is in question, generalization in CP-SFP is a difficult task to take up. More reliable CP-SFP models can be attained by applying deep learning, ensemble, and domain adaptation techniques. The explainable AI algorithms have provided software developers with confidence in the usage of predictive models.

2. Literature Review

The main idea of this work is to achieve more effective CPDP (Cross-Project Software Defect Prediction) with the help of federated meta-learning. The traditional CPDP techniques have problems of privacy and non uniform initiation. Federated learning enables more than two

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, ICETM 2025 , https://doi.org/10.63328/IJRDES-V7CIP2

businesses to collaboratively learn models without giving any source data. The study utilizes meta-learning, that enables quick optimization of the project. The proposed approach combines several modalities in order to minimise the cost of processing, enhance the reliability of failure prediction, and ensure the safety of data privacy. Saeed, M.S. (24): This page provides a thorough introduction into the various Ensems.

In order to become more skilled in the prediction of software failure in projects with the help of analytics, you must resolve the anomalies of the data. Machine learning systems actually perform worse when prediction is required on non-defective units not because dissimilarity between the defective units is insufficient.

Cost-sensitive learning, hybrid techniques, and resampling are some of the solutions to this problem. By applying the feature selection and transfer learning methods, a project may be interdependent thereby enhancing model generalization and obtaining correct predictions across various software environments. The complexities of projects, development processes, and coding standards are change sensitive and as such, the accuracy of predictions is influenced by the changes and thus the generalization is a difficult task in CP-SFP.

Reliable CP-SFP models can be achieved with the help of deep learning, the use of ensemble, and domain adaptation techniques. The fact that explainable AI algorithms have been included has provided more confidence to software developers using predictive models. It is also known that CP-SFP is able to enhance generalization and correct skewed data.

3. System Design and Models

Software quality assurance is a necessary part of software quality assurance because it helps predict which problems may be experienced before their execution. Conventional methods of training machine learning models are generally using data that is already part of the project.

The first process is the training of these models with the defect data of the project. Such data governance tactics can be effective to startups and small businesses, since they have less error records. In most cases, supervised learning paradigms, including decision trees, support vector machines and neural networks are applied to classify high risk modules of fault. However, the efficiency of these means depends on the amount and the quality of classified information. Anticipating flaws in all initiatives (CPDP) is one of the ways to deal with data concerns. This approach is founded on statistical data which indicates the mistakes of similar assignments.

The methods of domain adaptation and transfer learning are used to align the various project datasets in an attempt to make predictions more accurate. Nonetheless, the risk of changes in the project size, number of defects, and code patterns is alarming. Inefficient models due to skewed datasets complicate the process of making correct predictions in case the number of faulty modules is smaller than that of non-faulty modules. All classes which contain primacy are put first in these processes. The generalization in CPDP is always difficult because the models which are trained in one project tend to be ineffective in a different project. The need to develop models with a broad application across the dynamic sphere of software development is not an easy task. Some of the techniques that are used are ensemble learning, feature mapping and instance selection.

4. Proposed Methodology

The proposed technique the cross-project analytical framework enhances software failure prediction through data imbalance correction and the generalization. Our solution will guarantee consistency of feature regions in all projects by using advanced methods of transfer learning and not by using only the defect data in a single project. The strategy uses domain adaptation methods to take into consideration differences in software metrics and practice.

Consequently, models can be learnt with a broad variety of software development tasks. To increase the error detection accuracy, hybrid machine learning models integrate the use of both traditional and deep learning.

In order to deal with the problem of uneven data, the suggested approach combines advanced resampling methods, including SMOTE, with cost-sensitive learning. These techniques reduce bias in the classes and are trained using a set that consists of both malfunctioning and functioning modules. An ensemble learning method is utilized to increase the stability of the model and provide that the defect prediction performance is more than average in terms of project diversity.

This is a method in which several classifiers are integrated. The software metrics are given priority in the system involving the method of feature selection. As a result, the effectiveness of forecasting is increased and the expenses of calculations are reduced.

The improved generalization and adaptability of models to new situations can be provided by meta-learning and adaptive feature transformation. The system is active and is able to dynamically update the feature representation and decision-making abilities based on the new project data instead of depending on the model training. It is always very precise on diverse software

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, ICETM 2025 , https://doi.org/10.63328/IJRDES-V7CIP2

environments with a great dependency on the benchmark data sets in its validation methods.

The advent of explainable AI has enabled the provision of a solution, which could not be understood in the past. Software engineers may use these strategies to alert developers of any issues that may arise and that before they develop into serious problems.

Improved Management of Imbalanced Data - Sophisticated resampling methods and cost sensitive learning is used to improve the accuracy in fault predictions and reflect the minority class, which consist of faulty modules.

Better Generalization Over Projects - Transfer learning and domain adaptation can be used to make the model better at learning all the diverse sources. This will ensure that it is compatible with a wide range of applications.

Greater Accuracy of Fault Detection - The hybrid approach will improve the predictive accuracy and reduce the false positive/negative ratio since it combines conventional classifiers with deep learning and ensemble-based approaches.

Minimized Feature Discrepancies - Adaptive feature transformation and selection of features will help in the standardization of raw data, removal of discrepancies between projects and ensuring that the model does what it is supposed to do at all times.

Scalability and Continuous Learning - It can more easily cope with changing software production processes and improve them in case a continuous stream of new data is preserved.

5. Implementation

Service Tools: service provider You need to have an active password enabled account with the service provider in order to utilize this feature. Having passed the check-in procedure, he/she will be able to select training and assessment activities and access the data set. You were looking forward that these files will be found soon. Your next visit should include a bar chart indicating the efficacy of different testing and learning techniques and the rate of different software defect prediction.

The general precision of training and testing data, the nature and quantity of software defective prediction, all user profiles, etc can be accessed over the web. Manage and approve users This module has resulted into the manager being in a position to see a full list of all users. Administrators may assign/revoke rights using the name of a user, email address and physical address. remote user Before going ahead, ensure that the application form is fully filled. Once the registration is done, the information of the individual will be added to the database. After the registration process is completed, he/she shall be requested to give his/her login and his/her password. After the verification of the identity of a user, he/she can see his/her photo, estimate the risk of a software error and continue with the registration and authentication process.

6. Results

Figure.1 Prediction of Software Fault Type

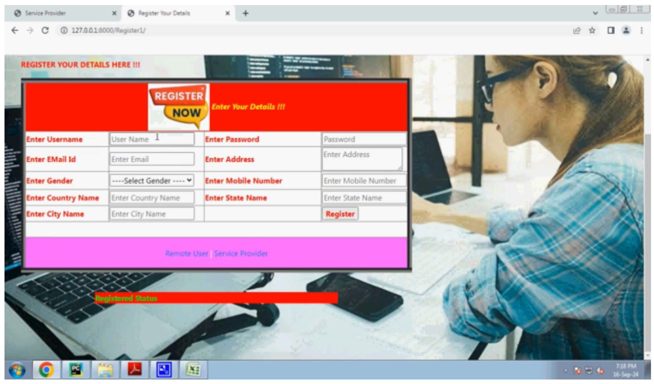

Figure.2 User Registration Page

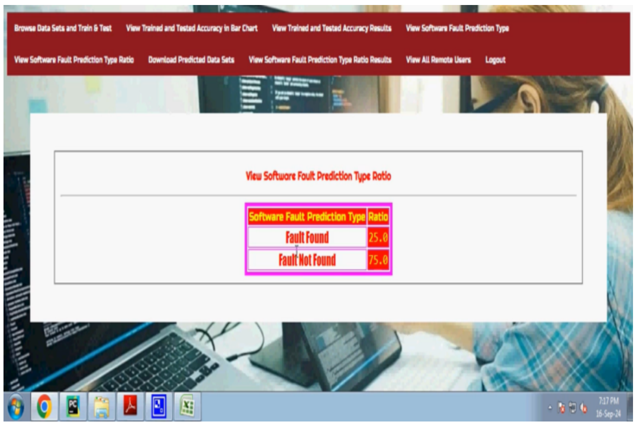

Figure.3 Software Fault Prediction Type Ratio

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, ICETM 2025 , https://doi.org/10.63328/IJRDES-V7CIP2

Figure. 4 Software Fault Prediction Type Details Page

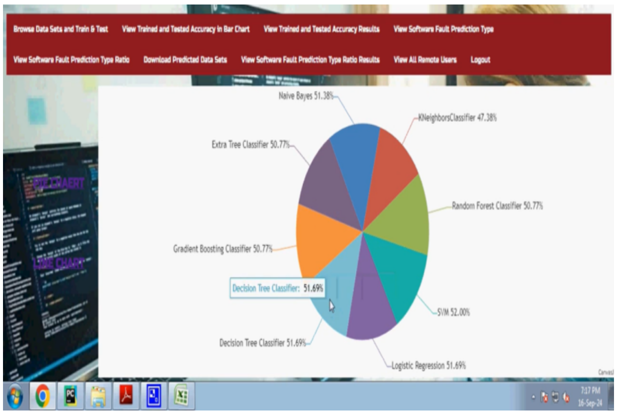

Figure.5 Software Fault Prediction in Pie Chart

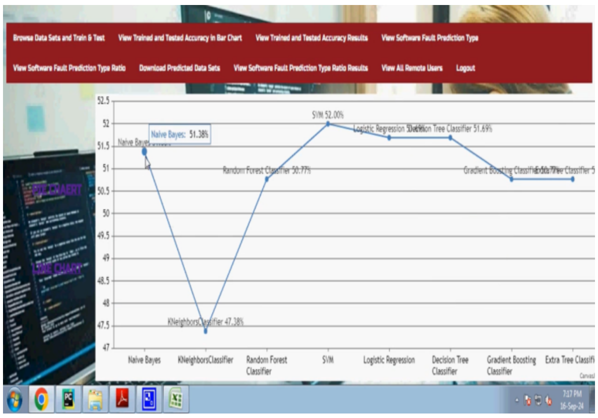

Figure. 6 Software Fault Prediction in Line Chart

Figure.7 Software Fault Prediction in Bar Chart

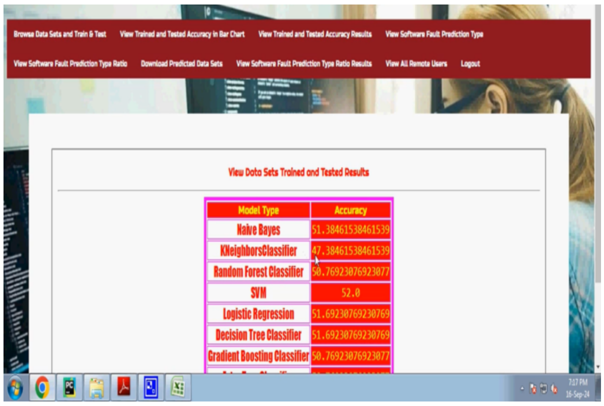

Figure. 8 Datasets Trained and Tested Results

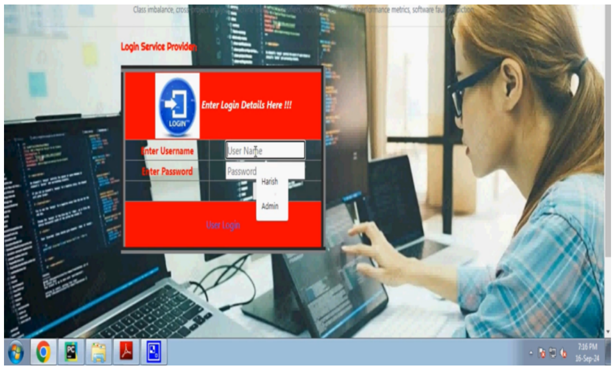

Figure.9 Service Provider Login Page

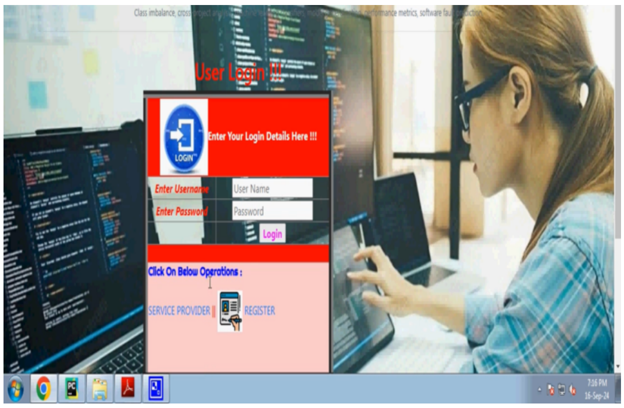

Figure. 10 User Login Page

7. Conclusion

Cross-project analysis may turn out to be useful regarding software quality, and, as a consequence, software defects forecasting. It is very useful when there is conflicting information or the problem of generalization. The suggested approach applies the latest machine learning models such as domain adaptation, transfer learning, and ensemble learning to address the problems that are inherent to the current project models. The integration of cost-sensitive learning and resampling methods will make sure that every error prone module is represented fairly. This way, there is less bias and accurate prediction.

ISSN: 2582 - 4201(Online) , http://www.ijrdes.com/ , IJRDES, Vol. 7, ICETM 2025 , https://doi.org/10.63328/IJRDES-V7CIP2

Adaptive feature transformation and meta-learning can be used to enhance the generalization of a model. It can forecast the software problems of the projects whose defects are distributed differently and whose code parameters vary. The system is capable of adjusting to the dynamic software development environment through analyzing and rectifying its previous mistakes. This solution applies interpretable AI to present key information to software engineers on key predictors of failures to enable them make more informed decisions. We also have a much more accurate prediction of software defects because of our approach, which tackles feature inconsistencies, classification biases and data missingness. The cross-project issues prediction is a necessary approach to ensure the software quality. As the precision and the efficiency of the fault detection increases, the software solutions will tend to be more stable, and the costs of the maintenance will decrease.

Reference

[1]. A. M. Buttar, U. Ahmad, A. H. Gumaei, A. Assiri, M. A. Akbar, and B. F. Alkhamees, "Deep Learning in Sign Language Recognition: A Hybrid Approach for the Recognition of Static and Dynamic Signs," *Mathematics 2023, Vol. 11, Page 3729*, vol. 11, no.17, p. 3729, Aug. 2023, doi: 10.3390/MATH11173729.

[2]. M. Madahana, K. Khoza-Shangase, N. Moroe, D. Mayombo, O. Nyandoro, and J. Ekoru, "A proposed artificial intelligence-based real-time speech-to-text to sign language translator for South African official languages for the COVID-19 era and beyond: In pursuit of solutions for the hearing impaired," *The South African Journal of Communication Disorders*, vol. 69, no. 2, p. 915, 2022, doi: 10.4102/SAJCD.V69I2.915.

[3]. Y. Zhang and X. Jiang, "Recent Advances on Deep Learning for Sign Language Recognition," *Computer Modeling in Engineering & Sciences*, vol. 139, no. 3, pp.2399–2450,Mar.2024 doi: 10.32604/CMES.2023.045731.

[4]. A. Baihan, A. I. Alutaibi, M. Alshehri, and S. K. Sharma, "Signlanguage recognition using modified deep learning network and hybrid optimization: a hybrid optimizer (HO) based optimized CNNSa- LSTM approach," *Scientific Reports 2024 14:1*, vol. 14, no. 1, pp. 1–22, Oct. 2024, doi: 10.1038/s41598-024-76174-7.

[5]. A. Sharma, D. Guo, A. Parmar, J. Ge, and H. Li, "Promoting Signlanguage Awareness: A Deep Learning Web Application for Signlanguage Recognition," *ACM International Conference Proceeding Series*, pp. 22–28, Nov. 2024, doi: 10.1145/3695719.3695723/ASSET/8EA1D820-783D-43F3-9B01-B63D8A32007A/ASSETS/GRAPHIC/ICDLT2024-4-FIG6.JPG.

[6]. K. Sahu, T. Saraswat, and R. Sawhney, "SignNet: A Deep Learning Architecture for Accurate Sign Language Recognition from Images," *2023 14th International Conference on Computing Communication and Networking Technologies,ICCCNT 2023*, 2023, doi: 10.1109/ICCCNT56998.2023.10306591.

[7]. M. Mahyoub, F. Natalia, S. Sudirman, and J. Mustafina, "Sign Language Recognition using Deep Learning," *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, vol. 2023-January, pp. 184–189, 2023, doi: 10.1109/DESE58274.2023.10100055.

[8]. F. M. Najib, "Sign language interpretation using machine learning and artificial intelligence," *Neural Comput Appl*, vol. 37, no. 2, pp. 841–857, Nov. 2024, doi: 10.1007/S00521-024-10395-9/TABLES/6.

[9]. M. S. Alam *et al.*, "ASL Champ!: A Virtual Reality Game with Deep-Learning Driven Sign Recognition," Dec. 2023, Accessed: Feb. 04, 2025. [Online]. Available: https://arxiv.org/abs/2401.00289v1

[10]. H. Orovwode, O. Ibukun, and J. A. Abubakar, "A machine learning-driven web application for sign language learning," *Front Artif Intell*, vol. 7, p. 1297347, Jun. 2024, doi: 10.3389/FRAI.2024.1297347/BIBTEX.

[11]. I. Papastratis, C. Chatzikonstantinou, D. Konstantinidis, K. Dimitropoulos, and P. Daras, "Artificial Intelligence Technologies for Sign Language," *Sensors 2021, Vol. 21, Page 5843*, vol. 21, no. 17, p. 5843, Aug. 2021,doi: 10.3390/S21175843.

[12]. A. Bailur, Y. Limbachia, M. Shah, H. Shah, and Prof. A. Kachare, "American Sign Language Recognition and its Conversion from Text to Speech," *Int J Res Appl Sci Eng Technol*, vol. 11, no. 9, pp. 1533–1538, Sep. 2023, doi: 10.22214/IJRASET.2023.55871.

[13]. H. Orovwode, O. Ibukun, and J. A. Abubakar, "A machine learning-driven web application for sign language learning," *Front Artif Intell*, vol. 7, p. 1297347, 2024, doi: 10.3389/FRAI.2024.1297347.

[14]. C. U. Bharathi, G. Ragavi, and K. Karthika, "Signtalk: Sign Language to Text and Speech Conversion," *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation, ICAECA 2021*, 2021, doi: 10.1109/ICAECA52838.2021.9675751.

[15]. V. G and K. Goyal, "Indian Sign Language Recognition Using Mediapipe Holistic," Apr. 2023, Accessed: Feb. 04, 2025. [Online]. Available: https://arxiv.org/abs/2304.10256v1