A Hybrid Nonlinear Manifold Detection Approach for Software Defect Prediction

Soumi Ghosh[1], Ajay Rana[2], Vineet Kansal[3]

[1,2]*Department of Computer Science & Engineering, Amity University Uttar Pradesh*
Noida, Uttar Pradesh, India
[1]*soumighoshphd@gmail.com;* [2]*ajay_rana@amity.edu*
[3]*Department of Computer Science & Engineering, Institute of Engineering and Technology Lucknow, India*
[3]*vineetkansal@yahoo.com*

Abstract: **Rapid development of software technology has influence on substantial industrial growth. Wide application of software in business related matters leads to development of reliable and defect free software system which is a challenging task. It requires development of effective techniques for prediction of software defects at early stage. For complexities in manual prediction of defects, automated techniques have come into effect. They are basically based on learning of pattern from earlier versions of software development and finding out the defects from the current version. Considerable impact of these techniques on industrial growth by predicting defects in software system attracted researchers in this field.**

In-spite of many studies performed by applying these techniques, desirable performance level and accurate defect prediction still remains a challenging task. For solving this problem, a hybrid technique based on Nonlinear Manifold Detection Techniques (Nonlinear MDTs) and machine learning for prediction of defects has been proposed in this paper. A new hybrid Nonlinear Manifold Detection (Nonlinear MD) Model has been applied for selecting and optimizing the features of software datasets that have been processed using Decision Tree (DT) and Random Forest (RF) classifications. Finally, a comparison and statistical evaluation of the experimental results obtained using new hybrid Nonlinear MD Model-DT have been made by Friedman test followed by Wilcoxon Sign rank test. The statistical outcome revealed that the proposed new hybrid Nonlinear MD Model-DT classification is better result oriented and more accurate in software defect prediction.

Keywords: **Decision Tree; Dimensionality Reduction; Feature Optimization; Friedman test; Nonlinear Manifold Detection; Software Defect Prediction; Wilcoxon sign rank test.**

I. INTRODUCTION

In the arena of software engineering, the testing phase is really a vital one, as it is connected with approaches like prediction and detection. Defect is a common occurrence in software system and proper prediction of those defects is a very challenging, tedious and complicated task due to the uncertainty inherent in it. It is also an expensive and time consuming exercise. Defects that arise due to human error in coding or computer programming may result in very poor quality of software product and thus sheer dissatisfaction of consumers.

For this crucial task of predicting or detecting defects in the software system, it is imperative to implement a defect management process which may reduce the defect density by predicting or locating the defective modules or defect prone areas. It is not at all possible to remove all the defects in the system but it is feasible to minimize the number of defects and thus improve the quality of software product. It depends on prediction of defects at very early stage of the system which can also reduce expenses and produce reliable and quality software.

In-fact, such defect prediction relates to a process for identification of errors in code of software with the help of application of machine learning or regression techniques [15]. These techniques help to extract information of previous software release and then identify the defects or errors in newer version. This process actually helps in evaluating only the faulty parts and not processing of entire current software code. As a result, development of good quality software is possible at minimized cost and lesser human effort. Therefore, defect prediction techniques have been proposed and adopted widely in various research studies with the idea of solving the problem of software defect prediction [18].

Various machine learning and data mining techniques have been used for predicting defects in software system [7] and classification technique is an integral part of machine learning methods which helps to categorize the software modules into defective or non-defective parts by examining the earlier instances. Different classification models generally used for classification of software datasets are statistical classification [4], neural network based classification [14], tree-based classification [11], [24] and analogy based classification [2]. P. Paramshetti and D. A. Phalke [16] used association rule mining for determining the frequency of detection of software defects inherent in the system. V. K. Dwivedi and M. K. Singh [26] put emphasis on developing a model based on data mining approaches for predicting defects in the early stage of software development life cycle in a much better way at reasonable expenses. I. H. Laradji *et al*. [9] developed a model for defect prediction by using ensemble based machine learning approach with selected features. A. Kaur [3] proposed a model based on different parameters of association rule mining and an improved apriori algorithm for prediction of software defects.

978-1-5386-4692-2/18/$31.00 ©2018 IEEE

A. Adline and M. Ramachandran [1] used data mining approach with genetic algorithm for improving quality development process of software by reducing both time complexity and cost effectiveness. X. Yang *et al.* [27] used learning to rank approaches on existing techniques for determining their accuracy level in case of software defect prediction. S. Ghosh *et al.* [21], [22] made an exhaustive review of research done in last two decades and found that Support Vector Machine, Neural Networks, Advance Machine Learning techniques have been widely used for defect prediction as those techniques were more effective and accurate compared to all others. T. Menzies *et al.* [25] also expressed similar views and considered that machine learning techniques are more effective in software defect prediction. Ghosh *et al.* [20] observed duly supported by statistical proof that Bayesian Network (BN) has much better performance and accuracy rate compared to other methods, when applied with or without MDTs.

Since last few decades many studies applying both machine learning and data mining have been undertaken for software defect prediction but none of them were totally effective and full proof. Though the use of machine learning techniques become more popular, but the problem of prediction of software defects was not at all resolved thoroughly [21].

Hence, in this paper, a new hybrid Nonlinear Manifold Detection (Nonlinear MD) Model, which is a hybrid technique based on Nonlinear Manifold Detection Techniques (Nonlinear MDTs) and machine learning methods, has been proposed for prediction of software defects.

II. CHALLENGING ISSUES WITH SOLUTIONS

Though series of research work have so far been made for software defect prediction at early stage [5], [6], [15], [21] but still many issues and challenges are still left in this field.

A. Feature Selection

A considerable number of feature selection techniques have been used for software defect prediction but it has been observed that mostly the selection of attributes are improper and having complexity as well as more time consuming and expensive. The techniques could not produce desired and accurate results. As such, it has become necessary to find out a new technique for selection of relevant features from software datasets.

B. Performance Measurements

There are no standard parameters and criteria for measuring the defect prediction performance of the techniques as well as comparison of performances of different modules. In case, performance of a technique has been measured using higher Recall then Precision level remains lower. On the other hand, when performance of technique has been measured using higher Precision level then Recall remains lower. So, for balancing the overall prediction performance of technique, selection of measures like higher AUC and F-Measure are required for accurate defect prediction.

C. Issues of Feature Selection Techniques in High Dimensional Datasets

The size and dimensions of the datasets and quality of the attributes or features contribute immensely in regard to desirable and effective performance of the techniques in software defect prediction. In most cases different feature selection methods have been used earlier to select only relevant features from the datasets but it leads to loss of data and its original properties. In order to overcome this problem, a new technique needs to be proposed and applied, which may select only relevant and most significant attributes without any loss or alteration of original properties of datasets by means of reduction of high-dimensional software datasets.

D. Framework related Problems

Various techniques have been applied by the researchers by using databases and software of different nature [4] and the process of operation is different in case of each single technique or software defect prediction datasets. Practically, no general framework is available, which may be suitably used for any or all types of software datasets for defect prediction.

Due to these reasons, we have proposed a new hybrid Nonlinear Manifold Detection (Nonlinear MD) Model, which is basically a hybrid technique based on Nonlinear Manifold Detection Techniques (Nonlinear MDTs) associated with classification methods for resolving these issues and improving the performance of classifiers in software defect prediction. The main objective of the proposed new model is prediction of defects by application of hybrid based different Nonlinear MDTs along with reduction and optimization of dimensions or features of datasets which is vital for accurate software defect prediction. This paper has been arranged as follows- Section III covers experimental context includes description of the proposed new model, experimental setup, parameter optimization and statistical methods. Section IV indicates detailed analysis of results, its comparison and statistical validation. Section V relates to final conclusion and future aspects.

III. EXPERIMENTAL CONTEXT

A. New Hybrid Nonlinear Manifold Detection Model

In this section, emphasis has been paid on resolving the existing issues and challenges by developing a new hybrid model for software defect prediction-Nonlinear Manifold Detection (Nonlinear MD) Model.

The proposed Nonlinear MD Model is developed basically to reduce the dimensions of software datasets from higher to lower one keeping intact the essential and basic properties of datasets. Nonlinear MD Model can effectively predict software defects by dimensionality reduction and optimized featured selection. There are two phases in the model: (i) Feature

Optimization by using hybrid technique based on Nonlinear Manifold Detection Techniques (Nonlinear MDTs), (ii) Machine Learning (DT and RF classification) techniques. The overall framework of the proposed model is presented in Fig 1.

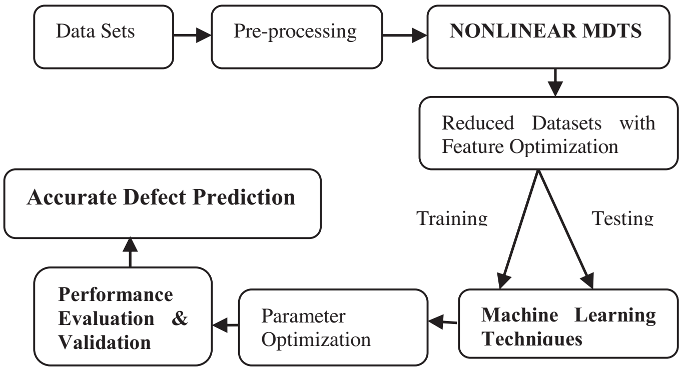

Fig. 1. Overall Framework of the Proposed Model

B. Nonlinear Manifold Detection Techniques for Feature Optimization

Proposed new hybrid Nonlinear MD Model is mainly based on different (eleven) Nonlinear MDTs, which reduce the higher dimensions of datasets into lower one more effectively in such a way that datasets contain minimum parameters which are pre-requisite for maintaining its original properties of datasets [10]. Nonlinear MDTs are very useful tool for classification, visualization and compression of datasets that are having high dimensions by lowering the computation and processing time and thus solving problems associated with complex, high-dimensional datasets [12]. A detailed description of different Nonlinear MDTs with their merits has already been given in earlier work [19], [20], [22], [23]. However, the proposed model uses eleven Nonlinear MDTs for feature optimization and selection, which have been briefly explained below.

ISOMAP (Isometric Feature Mapping) - It is a method nonlinear in character and applied widely for computation of embedding from high to low dimensional datasets.

LLE (Locally Linear Embedding) - This local nonlinear method is applied for reduction of dimension of datasets by preserving the original features of high-dimensional datasets.

LEM (Laplacian Eigen Maps) - The low dimensional datasets are computed retaining its properties and keeping the distance between a dataset and its nearest neighbors to minimal.

L-ISOMAP (Landmark ISOMAP) - It is an extended version of ISOMAP which performs faster than ISOMAP as data-points are meaningfully selected.

AE (AutoEncoders) - It aims at learning an encoding i.e. compressed representation of a dataset. It is obtained by using one of the many variants that are having back-propagation such as Conjugate method etc.

SPE (Stochastic Proximity Embedding) - It is basically a technique nonlinear in nature and applied for reduction of dimensions of datasets on the basis of neighborhood graph.

NPE (Neighborhood Preserving Embedding) - For reduction of dimensions of datasets, this method performs by minimizing the operational cost of a local nonlinear technique.

LPP (Linearity Preserving Projection) - It combines the benefits of local and nonlinear MDTs through a linear mapping for reducing the cost function of LEM technique.

SNE (Stochastic Neighborhood Embedding) – It is an iterative method similar to MDS and aims at retaining the pair-wise distance between data-points in case of low-dimensional representation of datasets.

Sym SNE (Symmetric Stochastic Neighborhood Embedding) - It is an extended and specialized version of SNE which mainly uses the Gaussian gradients for optimization of SNE.

Kernel PCA (Kernel Principal Component Analysis) -It is a kernel based method and an extended version of PCA. By applying a kernel, original functions of PCA is reproduced with a mapping of nonlinear nature.

C. Experimental Setup

The proposednew hybrid Nonlinear MD Model has been used and tested on four datasets-Camel,Arc,Ivy and Poi from open source software defect repository [13]. A brief details of datasets used has been given in Table I. The prediction performance of Decision Tree (DT) and Random Forest (RF) classification techniques has been measured using measures like Accuracy, F-measure and AUC.

TABLE I: Brief Details of Datasets Used

	Camel	**Arc**	**Ivy**	**Poi**
Version	1.6	1	2	3
Modules	965	234	352	442
Features	21	21	21	21
% Defective Modules	19.48	11.54	11.36	63.57

D. Parameter Optimization

The method used for optimization and validation also applied in reducing the bias of random sampling of datasets by dividing the datasets into n and n-1 parts of equal size with trained classifiers. The eliminated parts are applied in the test part of the datasets for evaluating the effectiveness of proposednew hybrid Nonlinear MD Model. Finally, evaluation of an average of performance of all n parts is done.

E. Statistical Methods

A nonparametric statistical test has been performed using Friedman Test and applied for finding out the significant difference in prediction performance of classification techniques along with proposed new hybrid Nonlinear MD Modeland thereafter ranking them in order [8].

Wilcoxon Sign rank test is a nonparametric test used for comparison of prediction performance of two classifiers across multiple domains [17].

IV. EXPERIMENT AND RESULTS WITH COMPARATIVE ANALYSIS

In this section, two experiments have been conducted on four datasets-Camel, Arc, Ivy and Poi. Initially, an experiment has been carried out without feature optimization where the prediction performance has been evaluated by using Decision Tree (DT) and Random Forest (RF) classification techniques. Another experiment has been performed using the proposed new Nonlinear MD Model, hybrid technique based on Nonlinear MDTs associated with DT and RF classification methods. Finally, a comparative analysis between the results of these two experiments have been made in terms of measures like Accuracy, F-measure, AUC and also statistically validated by using Friedman test followed by Wilcoxon Sign rank test.

A. Without Feature Optimization

In this subsection, the experimental analysis of four datasets-Camel,Arc,Ivy and Poi without feature optimization have been made. The prediction performance of DT and RF classification methodswithout feature optimization has been evaluated and optimized on these datasets using Accuracy and F-measure. The comparative analysis of prediction performance measurement of these two classification techniques on these four datasets have beenrepresented in Fig 2, 3, 4, 5.

B. With Feature Optimization using Proposed Nonlinear Manifold Detection Model

In this section, a hybrid technique based on Nonlinear MDTs associated with DT and RF classification methods; a new hybrid Nonlinear MD Model has been applied on Camel,Arc,Ivy and Poi datasets for selection as well as optimization of features. The initial step involves application of different Nonlinear MDTs like ISOMAP, LLE, LEM, L-ISOMAP, AE, SPE, NPE, LPP, SNE, Sym SNE and Kernel PCA on these high-dimensional datasets for reducing their dimensions by eliminating the unwanted, redundant and irrelevant features. The selected datasets with high dimensions have been reduced into lower three dimensional datasets, at a point where the embedding dimensions equals to two without inclusion of class labels. For application of DT and RF classification techniques, those new lower (three) dimensional datasets have been taken as inputs and they include relevant, most significant and optimized features. The prediction performance of DT and RF classification methodswithfeature optimization have been evaluated and optimized using cross-validation test on these datasets using Accuracy and F-measure. The comparative analysis of prediction performance measurement of these two classification techniques on these four datasets has beenrepresented in Fig 2, 3, 4, 5.

C. Results of Comparative Analysis

The outcome of the comparative analysis of prediction performance measurement of DT and RF classification methods obtained from these two above-mentioned experiments revealed that for Camel dataset, only DT classifier outperforms with new hybrid Nonlinear MD Model having accuracy rate of 80.518%. In respect of Poi dataset, DT classifier showed better accuracy with ISOMAP 77.602%, LLE 64.253%, LEM 78.054%, L-ISOMAP 75.792%, LPP 76.697%, Sym SNE 63.575% respectively. But, RF classifier performed well with some Nonlinear MDTs (AE, SPE, NPE, SNE and Kernel PCA) and also in case of without feature optimization with Accuracy (80.996%). For Ivy dataset, only DT classifier showed higher accuracy rate with (all Nonlinear MDTs) new hybrid Nonlinear MD Model. Simultaneously, DT showed better accuracy rate when used without feature optimization. Similarly, in Arc dataset, DT performed better with all Nonlinear MDTs and in case of without feature optimization except SPE.

The overall result of the comparative analysis showed that new hybrid technique based on Nonlinear MD Model-DT classification technique provides optimized features with better results and accuracy compared to all other techniques.

D. Result Validation

To validate the results of comparative analysis as to whether prediction performance measurement of new hybrid Nonlinear MD Model-DT classification technique is statistically different significantly or not than all other techniques, a statistical test-Friedman test has been performed. The critical value at level of significance £ (0.05) has been calculated at degree of freedom equals to one. It is found that the calculated value of X^2 is 3.841 and tabulated value X^2 is 3.930 for F-Measure by use of Chi-Square table. The prediction performance measurement of new hybrid Nonlinear MD Model-DT classification technique has been proved significant different than all other techniques because the P-value computedfor F-measure and AUC is 0.04743, 0.0001 respectively, less than £ (0.05). Further, performance measurement of classification techniques has been given rank for Friedman's rank computation. It is observed that the prediction performance measurement is better when lower is the mean rank. In Table II, the results of statistical-Friedman test based on mean ranking of all classification techniques for F-measure and AUC on all selected software datasets proved that proposed new hybrid Nonlinear MD Model-DT classification is significantly different and most accurate for software defect prediction.

Wilcoxon Sign rank test has been performed by assuming the null hypothesis that there is no effect or significant difference in prediction performance measurement of classification techniques due to new hybrid Nonlinear MD Model. Initially, for hypothesis testing, the confidence threshold value is equal to 0.05 (£). A box-plot with different colours have been represented in Fig 6 for more effective visualization that new hybrid Nonlinear MD Model-DT classification outperforms all other defect prediction models with statistical significance, as the calculated P-value from Wilcoxon Sign rank test for F-measure and AUC is 0.028, 0.0001 respectively, less than £. Fig 6 depicts that the degree of dispersion in the F-measure value of RF is relatively higher than DT classifier. As such, the mean difference between F-measure and AUC value of DT and RF is 0.06, 0.07 respectively. Hence, it has been validated from the statistical results of Wilcoxon Sign rank test given in Table III that new hybrid Nonlinear MD Model-DT classification performs best and accurate in software defect prediction.

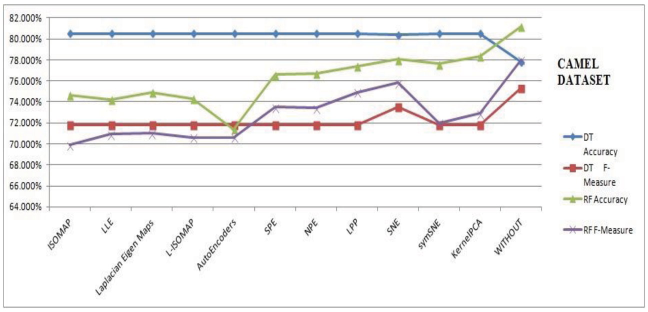

Fig. 2. The comparative analysis of prediction performance measurement of DT and RF classification techniques on Camel dataset

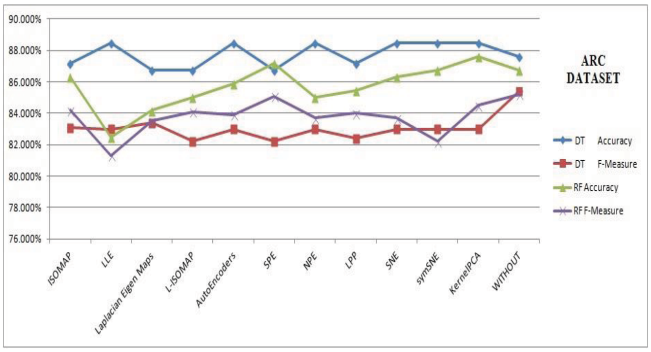

Fig. 3. The comparative analysis of prediction performance measurement of DT and RF classification techniques on Arc dataset

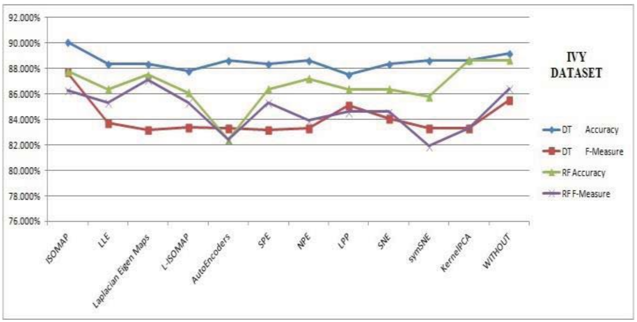

Fig. 4. The comparative analysis of prediction performance measurement of DT and RF classification techniques on Ivy dataset

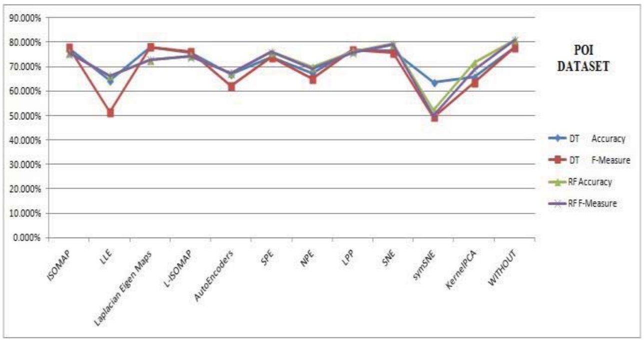

Fig. 5. The comparative analysis of prediction performance measurement of DT and RF classification techniques on Poi dataset

TABLE II: The Results of Friedman test based on Mean ranking of all classification techniques with new hybrid Nonlinear MD Model for F-measure and AUC

Classification Techniques	F-Measure Mean ranking	AUC Mean ranking	Mean	Std. Dev.
DT	1.352 (1)	1.114 (1)	0.767	0.086
RF	1.648 (2)	1.886 (2)	0.777	0.076

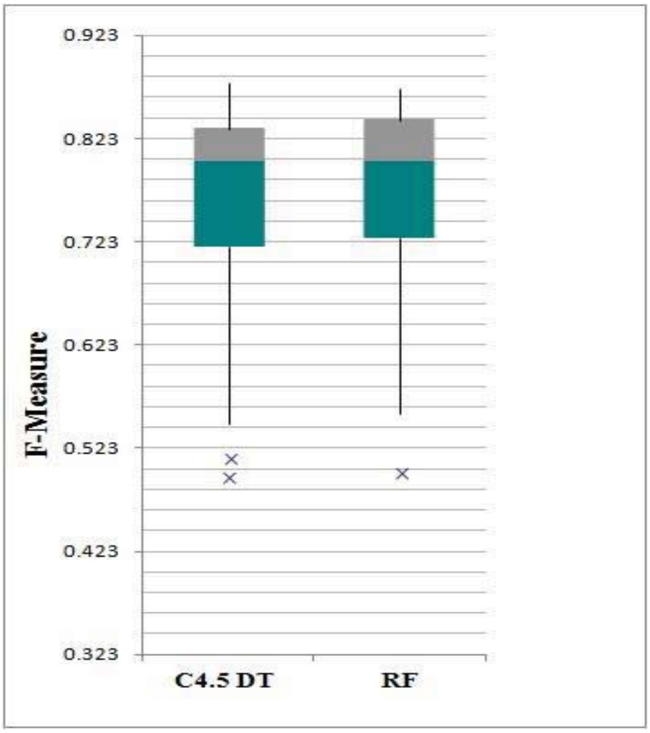

Fig. 6. A Box-Plot for comparative analysis of prediction performance measurement of DT and RF classification based on new hybrid Nonlinear MD Model in terms of F-Measure

TABLE III: Results of Wilcoxon Sign rank test for F-measure and AUC of all classification techniques with new hybrid Nonlinear MD Model

Parameters	F-Measure	AUC
Mean Diff. of DT and RF	0.06	0.07
P-value (Two-tailed)	0.028	0.0001
alpha	0.05	0.05

V. CONCLUSION

This research work focused and concentrated on development and application of a new hybrid Nonlinear Manifold Detection (Nonlinear MD) Model, which is pre-dominantly a hybrid technique based on Nonlinear Manifold Detection Techniques (Nonlinear MDTs) and machine learning for predicting software defects. Although, sizable number of techniques have so far been developed for early predicting defects in software but those techniques had their own drawbacks, limited effectiveness and classification accuracy for software defect prediction. For overcoming those challenging issues that have been identified from review of earlier research work, a new hybrid Nonlinear MDModel associated with classification techniques has been proposed for accurate prediction of defects with optimized feature selection strategy. In addition to this, comparison and statistical evaluation of experimental results have been made in an extensive manner by using tests like Friedman and Wilcoxon Sign rank test. It has been determined from the statistical results that new hybrid Nonlinear MD Model-DT classification achieved much better result and more accurate prediction performance than the existing techniques.

In future, the software developers may use this hybrid technique based on Nonlinear MDTs in an advanced manner for software defect prediction purpose using advance machine learning techniques for feature optimization with more accuracy. Wider application of this particular hybrid Nonlinear MD Model may enable the software developers to formulate quality projects for defect prediction with much higher accuracy at minimized cost and time.

REFERENCES

[1] A. Adline and M. Ramachandran, "Predicting the Software Fault using the Method of Genetic Algorithm," *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol.3, Issue 2, pp.390-398, 2014.

[2] A. Idri, F. A. Amazal and A. Abran, "Analogy-based Software Development Effort Estimation: A Systematic Mapping and Review," *Information and Software Technology*, vol. 58, pp. 206-230, 2015.

[3] A. Kaur, "Defect Prediction by Pruning Redundancy in Association Rule Mining," *International Journal of Advanced Research in Computer Science*, 2017.

[4] C. Seiffert, T. M. Khoshgoftaar, J. V. Hulse and A. Folleco, "An Empirical Study of the Classification Performance of Learners on Imbalanced and Noisy Software Quality Data," *Information Sciences*, vol. 259, pp. 571-595, 2014.

[5] C. Tantithamthavorn, S. McIntosh, A. E. Hassan and K. Matsumoto, "An Empirical Comparison of Model Validation Techniques for Defect Prediction Models," *IEEE Transactions on Software Engineering*, vol. 43, Issue 1, 2017.

[6] E. A. Felix and S. P. Lee, "Integrated Approach to Software Defect Prediction," *IEEE Access*, vol. 5, pp. 21524 – 21547, 2017.

[7] G. Mauša and T. G. Grbac, "Co-Evolutionary Multi-Population Genetic Programming for Classification in Software Defect Prediction: An Empirical Case Study," *Applied Soft Computing*, vol. 55, pp. 331-351, 2017.

[8] G. W. Corder and D. I. Foreman, *Nonparametric Statistics: A Step-by Step Approach*, 2nd ed., John Wiley & Sons, 2014.

[9] I.H. Laradji, Md. Alshayeb and L. Ghouti , "Software Defect Prediction using Ensemble Learning on Selected Features," *Information and Software Technology*, vol.58, pp.388-402, 2015. https://doi.org/10.1016/j.infsof.2014.07.005

[10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press Professional Inc., San Diego, CA, USA, 1990.

[11] L. Guo , Y. Ma , B. Cukic and H. Singh, "Robust Prediction of Fault-Proneness by Random Forests", in *Proc. 15th International Symposium on Software Reliability Engineering*, 2004.

[12] L. O. Jimenez and D. A. Landgrebe, "Supervised Classification in High-Dimensional Space: Geometrical, Statistical, and Asymptotical Properties of Multivariate Data," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 28, no. 1, pp. 39-54, 1997.

[13] M. Jureczko and L. Madeyski, "Towards identifying software project clusters with regard to defect prediction", in *Proc. 6th*

International Conference on Predictive Models in Software Engineering, Article No. 9, 2010.

[14] Ö. F. Arar and K. Ayan, "Software Defect Prediction using Cost-Sensitive Neural Network," *Applied Soft Computing*, vol. 33, pp. 263-277, 2015.

[15] P. Bishnu and V. Bhattacherjee, "Software Fault Prediction using Quad Tree based K-Means Clustering Algorithm*," IEEE Transactions on Knowledge and data Engineering*, vol. 24, no. 6, pp. 1146-1150, 2012.

[16] P. Paramshetti and D. A. Phalke, "Software Defect Prediction for Quality Improvement using Hybrid Approach,"*International Journal of Application or Innovation in Engineering & Management*, vol.4, no.6, pp.99-104, 2015.

[17] R. Lowry, Concepts and Applications of Inferential Statistics, 2011.

[18] R. Malhotra, "A Systematic Review of machine learning Techniques for Software Fault Prediction," *Applied Soft Computing*, vol. 27, pp. 504-518, 2015.

[19] S. Ghosh, A. Rana and V. Kansal, "A Nonlinear Manifold Detection based Model for Software Defect Prediction", in *Proc. International Computational Intelligence and Data Science" ICCIDS-2018, Procedia Computer Science*, vol. 132, pp. 581-594, 2018.

[20] S. Ghosh, A. Rana and V. Kansal, "A Statistical Comparison for Evaluating the Effectiveness of Linear and Nonlinear Manifold Detection Techniques for Software Defect Prediction," *International Journal of Advanced Intelligence Paradigms*, Accepted, 2017.

[21] S. Ghosh, A. Rana and V. Kansal, "Predicting Defect of Software System", in *Proc. 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing*, Springer, Singapore, vol. 516, pp. 55- 67, 2017.

[22] S. Ghosh, A. Rana and V. Kansal, "Software Defect Prediction System Based on Linear and Nonlinear Manifold Detection", in *Proc. 4th International Conference on "Computing for Sustainable Global Development" INDIACom-2017*, Accepted, 2017.

[23] S. Ghosh, A. Rana and V. Kansal, "Statistical Assessment of Nonlinear Manifold Detection-based Software Defect Prediction Techniques," *International Journal of Intelligent Systems Technologies and Applications*, Accepted, 2018.

[24] S. S. Rathore and S. Kumar, "A Decision Tree Logic based Recommendation System to Select Software Fault Prediction Techniques," *Computing*, vol. 99, Issue 3, pp. 255–285, 2017.

[25] T. Menzies, J. Greenwald, and A. Frank, "Data Mining Static Code Attributes to Learn Defect Predictors," *IEEE Transactions on Software Engineering*, vol. 32, no. 11, pp. 2–13, 2007.

[26] V. K. Dwivedi and M. K. Singh, "Software Defect Prediction using Data Mining Classification Approach," *International Journal of Technical Research and Applications*, vol.4, no.6, pp.31-35, 2016.

[27] X.Yang, K.Tang and X.Yao, "A Learning-to-Rank Approach to Software Defect Prediction," *IEEE Transactions on Reliability*, vol.64, no.1, pp.234-246, 2015. https://doi.org/10.1109/TR.2014.2370891.