



Accuracy Comparison of Empirical Studies on Software Product Maintainability Prediction

Sara Elmidaoui, Laila Cheikhi^(✉), and Ali Idri

Software Project Management Research Team, ENSIAS,
Mohammed V University, Rabat, Morocco

Sara.Elmidaoui@gmail.com, Laila.Cheikhi@gmail.com,
Idri.Ali123@gmail.com

Abstract. Software maintainability is a very broad activity which ensures that the software product fulfills its changing requirements and enhancement capabilities once on the client side. Predicting software product maintainability contributes to the reduction of software product maintenance costs. In this perspective, many software product maintainability prediction (SPMP) techniques have been proposed in the literature. Some studies have empirically validated their proposed techniques while others have compared the accuracy of the SPMP techniques. This paper reviews a set of 29 studies, which are identified from eight digital libraries and collected from 2000 to 2017. The present paper is targeted to present the various SPMP techniques used and reveals all about the experimental design of these studies.

Keywords: Software product maintainability · Prediction techniques
Accuracy criteria

1 Introduction

Currently the software is ubiquitous in organizations whatever their sectors of activities, and its quality is becoming a critical issue. The main objective is to achieve the expressed and implicit needs (expectations), defined by the users when the software is used under specified conditions. A quality model has been defined in ISO 25010 standard which classifies and structures all software product quality characteristics/attributes. Maintainability as a quality characteristic for software product is related to the ease with which the software undergoes changing requirements [1].

Several research studies have been published related to software product maintainability prediction techniques (SPMP). Some studies have empirically validated their proposed techniques such as in [2, 3], others have conducted SLRs in order to have an overview of the existing studies in the literature such as in [4–7], while other studies have empirically compared their proposed SPMP techniques with some existing ones, such as in [8–10]. However, to the best of our knowledge no study has been published with the purpose of reviewing and analyzing these SPMP comparative studies based on their accuracy.

This paper analyzes a set of 29 empirical studies published between 2000–2017, that have proposed and/or evaluated and compared different SPMP techniques. This set of studies is identified from 8 digital libraries (Science Direct, Springer Link, Ebsco, ACM Digital Library, Google Scholar, Scopus and Jstore). Since each of these SPMP techniques was built in a particular experimental context, this paper review and analysis are addressed according to five 5 research questions (RQs) related to: SPMP techniques, metrics, datasets, accuracy criteria and validation procedures.

This paper is structured as follows. Section 2 presents the method used to collect the studies. Section 3 analyzes and discusses the review results and provides responses to the 5RQs. Section 4 provides an accuracy comparison of SPMP techniques reported superior in the literature. Conclusion and future work are presented in Sect. 5.

2 Method

This paper provides the results of the accuracy comparison of software product maintainability prediction techniques based on literature review. A systematic review was done through 8 databases; Science Direct, Springer Link, Ebsco, ACM Digital Library, Google Scholar, Scopus and Jstore. The search string was build using the following keywords (maintainability OR analyzability OR modifiability OR testability OR portability) AND (compar* OR empirical* OR evaluation* OR validation* OR experiment* OR control experiment OR case study OR survey) AND (software* OR application OR system) AND (predict* OR evaluat* OR assess* OR estimat* OR measur*) AND (method* OR technique*). This combination of keywords doesn't much in some databases, it was therefore necessary to adapt the search string for each database.

To reduce the large number of papers returned, the search was limited to papers published between 2000 and 2017. To get relevant papers, some inclusion and exclusion criteria were applied. Included papers were those conducting empirical studies on SPMP and/or providing comparison of SPMP techniques. Papers which fall outside the topic of maintainability, not including empirical comparison, or written in other languages than English were excluded. This task was performed by two authors. The conclusion to include or exclude a paper was done based on reading the abstracts in first time, and then reading the full papers. As a result of this search process, a set of 29 relevant papers was selected; grouped by years of publications in Table 1.

Table 1. Selected studies per years of publications

Years of publications	Studies
From 2003 to 2009	[8–14]
From 2010 to 2013	[15–27]
From 2014 to 2017	[28–36]

This set of 29 selected papers was used to provide responses to the following 5 Research Questions (RQs):

Acronyms: Ward Neural Network (WNN), General Regression Neural Network (GRNN), Bayesian Network (BN), Regression Tree (RT), Backward Elimination (BE), Stepwise Selection (SS), Multiple Adaptive Regression Splines (MARS), Multiple Linear Regression (MLR), Support Vector Regression (SVR), Gaussian Mixture Model (GMM), Aggregating One-Dependence Estimators (AODE), k Nearest Neighbor (KNN), Naïve Bayes (NB), Random Forest (RF), Radial Basis Function Network (RBF), Projection Pursuit Regression (PPR), Feed Forward Neural Network (FFNN), Fuzzy Inference Systems (FIS), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), Extreme Learning Machine (ELM), Multilayer Perceptron (MLP), Group Method of Data Handling (GMDH), Genetic Algorithms (GA), Probabilistic Neural Network (PNN), Linear Regression (LR), Multiple Classifiers Combination (MCC), Back Propagation Neural Network (BPNN), Mamdani Fuzzy Logic (MFL), Sensitivity Based Linear Learning Method (SBLLM), Least Median of Squares Regression (LMSR), Reduced Error Pruned Tree (REPTree), Locally Weighted Learning (LWL), Conjunctive Rule Learner (CR), Decision Table (DTable), M5 Rules (M5R), Pace Regression (PR), Isotonic Regression (IR), Regression By Discretization (RegByDisc), Additive Regression (AR), Ensemble Selection (ES), Gaussian Process Regression (GPR), Fuzzy Subtractive Clustering (FSC), Decision Stump (DS), K-means clustering (KMC), X-means clustering (XMC), Feed Forward 3-Layer Back Propagation Network (FF3LBPN), Logistic Regression (LogR), Kohonen Network (KN), Gene Expression Programming (GEP), Hybrid approach of neural network and genetic algorithm (Neuro-GA), Type-2 fuzzy logic systems (T2FLS), Hybrid approach of functional link artificial neural network with GA (FGA, AFGA), Particle Swarm Optimization with GA (FPSO, MFPSO), Clonal Selection Algorithm with GA (FCSA), SVR with Radial Kernel Function (SVR-RK).

Figure 1 depicts the classification of SPMP techniques used in the 29 studies. The analysis shows that the machine learning techniques were the most used with 83% (24 papers) compared to statistical techniques with 41% (12 papers). Note that some studies have evaluated many techniques, while others have focused on only one technique. Within the machine learning categories, the NN techniques were used in 16 studies, SVM in 7 studies, DT in 5 studies, EA in 5 studies, NF in 4 studies, BL in 3 studies, IB and CL in 2 studies each, and RBL, EL, and IR in 1 study each. The statistical category includes techniques such as: PR, MLR, MARS, RegByDisc, GMM, Lreg, SS, BE, LR, PPR, SVR, AR, GPR, SVR-RK, RT, LMSR, IR. The MLR technique was the most used (4 studies), followed by MARS (3 studies), then SS, BE, PPR, LogR, SVR, and RT (2 studies each).

3.2 Datasets Used for SPMP (RQ2)

Several datasets have been used in the 29 SPMP selected studies. Figure 2 reports the number of studies per datasets. According to the figure, UIMS and QUES were the most commonly used in 19 and 18 studies respectively; these datasets are provided by Li and Henry [40]. FLMS (File Letter Monitoring System), EASY (Classes Online Services collection), Lucene (Apache Lucene), JEdit (Java text Editor), and MIS (Medical Imaging System) are open source datasets (OSS) and have been used in two studies each. The remained datasets have been used only once.

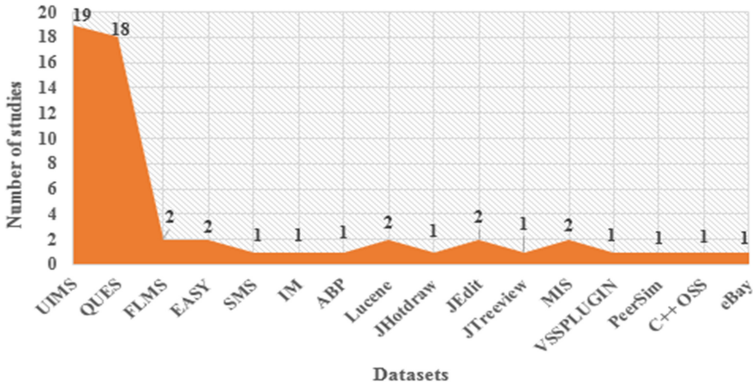


Fig. 2. Number of studies per datasets

3.3 Types of Metrics Used for SPMP (RQ3)

Many metrics have been proposed in the literature in order to evaluate the software product design as well as source code. The most used types of metrics in the 29 selected SPMP studies are presented in Table 2. As shown in the table, the C&K and Size metrics are the most used metrics (90%, respectively) compared to L&H, McCabe, MI, and Halstead’s metrics. It should be noted that the selected studies have used one or several metrics depending on the study purpose. For instance, studies [22, 28, 29] have used C&K, L&H, and Size metrics, while studies [11, 12] have used Halstead’s, McCabe, and Size metrics.

Table 2. Distribution of metrics types per studies

Metrics type	# of studies	Percent
Chidamber and Kemerer (C&K)	26	90%
Size	26	90%
Li and Henry (L&H)	23	79%
McCabe Cyclomatic Complexity	4	14%
Maintainability Index (MI)	1	3%
Halstead’s	1	3%
Miscellaneous	7	24%

Moreover, within the set of OO Metrics (OOM), it is noticed that 9 out of the 10 have been frequently used in the selected studies. Figure 3 presents the distribution of studies per these OOM. As shown, the Lack of Cohesion in Methods (LCOM), Depth of Inheritance Tree (DIT), Number of Children (NOC), Weighted Methods per Class (WMC), Response For a Class (RFC), Message Passing Coupling (MPC), Data Abstraction Coupling (DAC), Number Of local Methods (NOM), and Size have been used in more than 20 studies while Coupling Between Object (CBO) was used only in six studies.

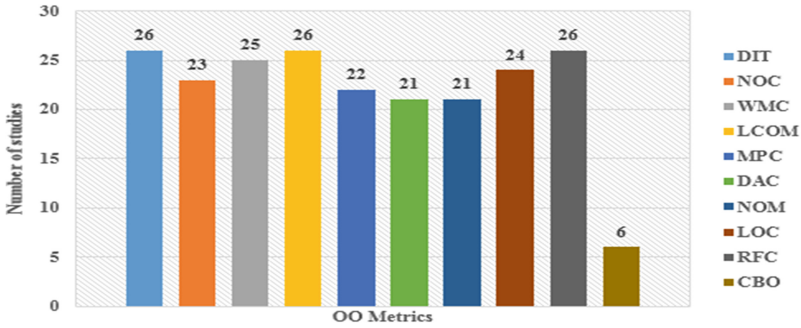


Fig. 3. Number of OO metrics per study

3.4 Accuracy Criteria and Validation Procedures Used for SPMP (RQ4)

In the 29 selected studies, various accuracy criteria were used to compare the accuracy of SPMP techniques (see Fig. 4). From this figure, it appear that MMRE is the most frequently used accuracy criteria; being adopted by 52% (15 papers), followed by the Pred (Pred (0.25), Pred (0.30), and Pred (0.75)) with 45% (13 papers), and Max MRE with 28% (8 papers).

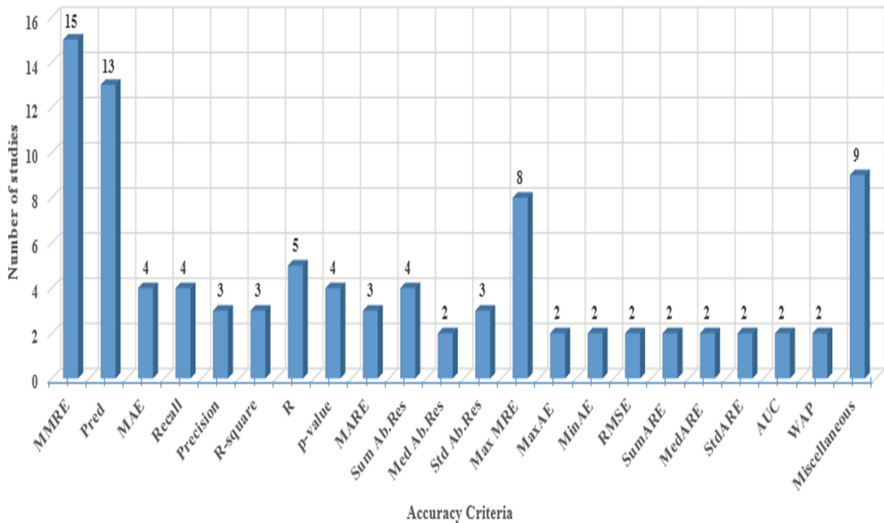


Fig. 4. Number of studies per accuracy criteria

Besides, in order to evaluate the accuracy of their proposed SPMP techniques, the selected studies have used different validation methods. Figure 5 reports the cross validation methods (CV) per studies. According to the figure, it is observed that k-FCV (k-Fold Cross-Validation) especially; 10-FCV method was the most used in the selected studies followed by LOOCV (Leave-One-Out Cross-Validation), CV, and at last 5 FCV.

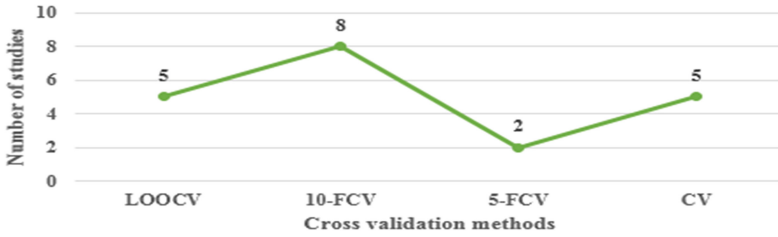


Fig. 5. Cross validation methods per studies.

3.5 SPMP Techniques Reported Superior from Literature (RQ5)

The 29 selected studies investigated in this paper have compared the accuracy of some SPMP techniques (new models and/or already published models). From the literature results, we can notice that no technique is definitively better than the others. For instance:

- MLP technique was reported superior in 4 studies,
- SVM, DT, ELM, GMDH and MARS techniques were reported superior in 2 studies each, and
- the rest of techniques are reported only once.

Moreover, the MARS technique was reported to be superior in two studies [10, 36], while not in 9 studies [13, 14, 18, 23, 25, 32–35], and SVM technique was reported accurate in 2 studies [21, 31] and not in [11, 12, 23, 35, 36]. Therefore, the choice of the best technique to predict the maintainability is not obvious since every technique has its advantages and its drawbacks.

4 SPMP Techniques Accuracy Comparison

The purpose of this section is to compare the techniques reported to be superior and which have used the same datasets (UIMS and QUES), metrics (L&H and C&K), accuracy criteria (MMRE, Pred (0.25), Pred (0.30)), and OO software applications. From the 29 selected studies, 7 studies (8 experiments) have been selected for UIMS set and 7 studies (9 experiments) have been selected for the QUES. Using the MMRE and the Pred as accuracy criteria for comparison, it is important to note that low MMRE value or a high Pred (25) value indicates good accuracy [41, 42].

Table 3 shows that the T2FLS has achieved a significantly better prediction accuracy than the other techniques with an MMRE value of 0.00007, Pred (0.25) value of 0.86 and Pred (0.30) value of 0.92 in UIMS datasets. Therefore, the T2FLS technique can predict maintainability of the UIMS dataset better than BN, MARS, TreeNet, ELM, MFL, FSC, and K*.

For QUES dataset, Table 4 shows that MFL, K*, and KNN have achieved the same MMRE value of 0.27. Moreover, they are slightly equal in terms of Pred: (Pred (0.25) = 0.52 and Pred (0.30) = 0.62 for MFL, Pred (0.25) = 0.62 and Pred (0.30) = 0.65 for

Table 3. Prediction accuracy for UIMS

Source	Technique	MMRE	Pred (0.25)	Pred (0.30)
[9]	BN	0.97	0.44	0.46
[10]	MARS	1.86	0.28	0.28
[14]	TreeNet	1.57	0.31	0.41
[16]	ELM	0.96	0.39	0.45
[23]	MFL	0.53	0.30	0.35
[25]	FSC	0.65	0.33	0.41
[25]	K*	0.56	0.36	0.41
[33]	T2FLS	0.00007	0.86	0.92

Table 4. Prediction accuracy for QUES

Source	Technique	MMRE	Pred (0.25)	Pred (0.30)
[9]	BN	0.45	0.39	0.43
[10]	MARS	0.32	0.48	0.59
[14]	TreeNet	0.42	0.58	0.65
[16]	ELM	0.35	0.36	0.38
[23]	MFL	0.27	0.52	0.62
[24]	SBLLM	0.34	0.50	0.56
[25]	PR	0.37	0.39	0.52
[25]	K*	0.27	0.56	0.66
[25]	KNN	0.27	0.62	0.65

KNN and Pred (0.25) = 0.56 and Pred (0.30) = 0.66 for K*). Thus, overall, the MFL, K* and KNN techniques are better than BN, MARS, TreeNet, ELM, SBLLM, and PR.

5 Conclusion and Future Work

In this paper, we have investigated the empirical studies that have compared the accuracy of some SPMP models, published between 2000 and 2017. The results of the discussion conducted show that the NN, SVM, DT, EA, and MLR techniques and the DIT, LCOM, RFC, and WMC measures were commonly used to predict maintainability. The UIMS and QUES datasets, the MMRE and Pred(20or30) accuracy criteria, and the 10-fold cross validation method were the frequently used in the selected studies. The results have also showed that there is no accurate technique. As a future work, it is planned to conduct more empirical studies to better predict the maintainability of software.

References

1. ISO. Systems and Software engineering — Systems and Software Quality Requirements and Evaluation — System and Software Quality Models. ISO/IEC 25010, 34 p. International Organization for Standardization, Geneva, Switzerland (2010)
2. Dhankhar, P., Mittal, H.K., Mittal, A., et al.: Maintainability prediction for object oriented software. *Int. J. Adv. Eng. Sci.* **1**(1), 8–11 (2011)
3. Kiewkanya, M., Jindasawat, N., Muenchaisri, P.: A methodology for constructing maintainability model of object-oriented design. In: *Proceedings of the Fourth International Conference on Quality Software, QSIC 2004*, pp. 206–213. IEEE (2004)
4. Riaz, M., Mendes, E., Tempero, E.: A systematic review of software maintainability prediction and metrics. In: *Proceedings of the 2009 3rd International Symposium on Empirical Software Engineering and Measurement*, pp. 367–377. IEEE Computer Society (2009)

5. Riaz, M.: Maintainability prediction of relational database-driven applications: a systematic review. In: 16th International Conference on Evaluation Assessment in Software Engineering (EASE), pp. 263–272 (2012)
6. Elmidaoui, S., Cheikhi, L., Idri, A.: A survey of empirical studies in software product maintainability prediction models. In: 11th International Conference on Intelligent Systems: Theories and Applications (SITA), pp. 1–6. IEEE (2016)
7. Elmidaoui, S., Cheikhi, L., Idri, A.: Software product maintainability prediction: a survey of secondary studies. In: 4th International Conference on Control, Decision and Information Technologies (CoDIT) (2017)
8. Thwin, M.M.T., Quah, T.S.: Application of neural networks for estimating software maintainability using object-oriented metrics. In: SEKE, pp. 69–73 (2003)
9. Van Koten, C., Gray, A.R.: An application of Bayesian network for predicting object-oriented software maintainability. *Inf. Softw. Technol.* **48**(1), 59–67 (2005)
10. Zhou, Y., Leung, H.: Predicting object-oriented software maintainability using multivariate adaptive regression splines. *J. Syst. Softw.* **80**(8), 1349–1361 (2006)
11. Jin, X., Liu, Y., Ren, J., Xu, A., Bie, R.: Locality preserving projection on source code metrics for improved software maintainability. In: Proceedings of AI 2006, pp. 877–886 (2006)
12. Tian, Y., Chen, C., Zhang, C.: AODE for source code metrics for improved software maintainability. In: Fourth International Conference on Semantics, Knowledge and Grid, SKG 2008, pp. 330–335. IEEE (2008)
13. Li-Jin, W., Xin-Xin, H., Zheng-Yuan, N., Wen-Hua, K.: Predicting object-oriented software maintainability using projection pursuit regression. In: 2009 1st International Conference on Information Science and Engineering (ICISE), pp. 3827–3830 (2009)
14. Elish, M.O., Elish, K.O.: Application of TreeNet in predicting object-oriented software maintainability: a comparative study, pp. 69–78 (2009)
15. Kaur, A., Kaur, K., Malhotra, R.: Soft computing approaches for prediction of software maintenance effort. *Int. J. Comput. Appl.* **1**(16), 69–75 (2010)
16. Olatunji, S.O., Rasheed, Z., Sattar, K.A., Al-Mana, A.M., Alshayeb, M., El-Sebakhy, E.A.: Extreme learning machine as maintainability prediction model for object oriented software systems. *J. Comput.* **2**(8), 42–56 (2010)
17. Dubey, S.K., Rana, A., Dash, Y.: Maintainability prediction of object-oriented software system by multilayer perceptron model. *SIGSOFT* **37**(5), 1–4 (2012)
18. Malhotra, R., Chug, A.: Software maintainability prediction using machine learning algorithms. *Int. J. (SEIJ)* **2**(2), 19–36 (2012)
19. Dash, Y., Dubey, S.L., Rana, A.: Maintainability prediction of object oriented software system by using artificial neural network approach. *Int. J. Soft Comput. Eng. (IJSCE)* **2**(2), 420–423 (2012)
20. Hegedus, P., Ladanyi, G., Siket, I., et al.: Towards building method level maintainability models based on expert evaluations. In: Computer Applications for Software Engineering, Disaster Recovery, and Business Continuity, pp. 146–154. Springer, Heidelberg (2012)
21. Aljamaan, H., Elish, M.O., Ahmad, I.: An ensemble of computational intelligence models for software maintenance effort prediction. In: International Work-Conference on Artificial Neural Networks, pp. 592–603. Springer, Heidelberg (2013)
22. Ye, F., Zhu, X., Wang, Y.: A new software maintainability evaluation model based on multiple classifiers combination. In: International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (QR2MSE), pp. 1588–1591 (2013)
23. Ahmed, M.A., Al-Jamimi, H.A.: Machine learning approaches for predicting software maintainability: a fuzzy-based transparent model. *IET Softw.* **7**(6), 317–326 (2013)

24. Olatunji, S.O., Ajasin, A.: Sensitivity-based linear learning method and extreme learning machines compared for software maintainability prediction of object-oriented software systems. *ICTACT J. Soft Comput.* **3**, 514–523 (2013)
25. Kaur, A., Kaur, K.: Statistical comparison of modeling methods for software maintainability prediction. *Int. J. Softw. Eng. Knowl. Eng.* **23**, 743–774 (2013)
26. Mehra, A., Dubey, S.K.: Maintainability evaluation of object-oriented software system using clustering techniques. *IJCT* **5**(2), 136–143 (2013)
27. Malhotra, R., Chug, A.: Application of group method of data handling model for software maintainability prediction using object oriented systems. *Int. J. Syst. Assur. Eng. Manag.* **5**(2), 165–173 (2014)
28. Kaur, A., Kaur, K., Pathak, K.: A proposed new model for maintainability index of open source software. In: 3rd International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), pp. 1–6 (2014)
29. Kaur, A., Kaur, K., Pathak, K.: Software maintainability prediction by data mining of software code metrics. In: 2014 International Conference on Data Mining and Intelligent Computing (ICDMIC), pp. 1–6. IEEE (2014)
30. Malhotra, R., Chug, A.: A metric suite for predicting software maintainability in data intensive applications. In: Transactions on Engineering Technologies: Special Issue of the World Congress on Engineering and Computer Science, pp. 161–175 (2014)
31. Elish, M.O., Aljamaan, H., Ahmad, I.: Three empirical studies on predicting software maintainability using ensemble methods. *Soft Comput.* **19**(9), 2511–2524 (2015)
32. Kumar, L., Rath, S.K.: Neuro – genetic approach for predicting maintainability using Chidamber and Kemerer software metrics suite. In: Recent Advances in Information and Communication Technology, pp. 31–40. Springer, Cham (2015)
33. Olatunji, S.O., Selamat, A.: Type-2 fuzzy logic based prediction model of object oriented software maintainability. *Communications in Computer and Information Science*, vol. 513, pp. 329–342 (2015)
34. Soni, A.K., Lobiyal, D.K., Kumar, L., Naik, D.K., Rath, S.K.: Validating the effectiveness of object-oriented metrics for predicting maintainability. In: 3rd International Conference on Recent Trends in Computing (ICRTC 2015). *Procedia Computer Science*, vol. 57, pp. 798–806 (2015)
35. Kumar, L., Rath, S.K.: Hybrid functional link artificial neural network approach for predicting maintainability of object-oriented software. *J. Syst. Softw.* **121**, 170–190 (2016)
36. Kumar, L., Rath, S.K., Sureka, A.: Using source code metrics and multivariate adaptive regression splines to predict maintainability of service oriented software. In: International Symposium on High Assurance Systems Engineering (HASE), Singapore (2017)
37. Idri, A., Amzal, F.A., Abran, A.: Analogy-based software development effort estimation: a systematic mapping and review. *Inf. Softw. Technol.* **58**, 206–230 (2015)
38. Wen, J., Li, S., Lin, Z., Hu, Y., Huang, C.: Systematic literature review of machine learning based software development effort estimation models. *Inf. Softw. Technol.* **54**, 41–59 (2012)
39. Malhotra, R.: A systematic review of machine learning techniques for software fault prediction. *Appl. Soft Comput.* **27**, 504–518 (2015)
40. Li, W., Henry, S.: Object-oriented metrics that predict maintainability. *J. Syst. Softw.* **23**(2), 111–122 (1993)
41. Conte, S.D., Dunsmore, H.E., Shen, V.Y.: *Software Engineering Metrics and Models*. Benjamin Cummings Publishing Co., Inc., Redwood City (1986)
42. Gray, A.R., MacDonell, S.G.: A comparison of techniques for developing predictive models of software metrics. *Inf. Softw. Technol.* **39**(6), 425–437 (1997)