Statistical Analysis of the Three-Dimensional Data of Software Metrics RFC, CBO, and WMC that are not Normally Distributed

Sergiy Prykhodko[1,2], Lidiia Makarova[1] and Andrii Pukhalevych[1]
[1]*Department of Software for Automated Systems, Admiral Makarov National University of Shipbuilding, Heroes of Ukraine Aveneu 9, 54007 Mykolaiv, Ukraine*
[2]*Department of Software Engineering, Odesa Polytechnic National University, Shevchenko Avenue 1, 65044 Odesa, Ukraine*
sergiy.prykhodko@nuos.edu.ua, lidiia.makarova@nuos.edu.ua, andrii.pukhalevych@nuos.edu.ua

Keywords: Software Quality, Software Development, Statistical Analysis, Prediction Interval, Confidence Interval, Nonlinear Regression, OOD Metric, RFC, CBO, WMC, Multivariate Box-Cox Transformation, Outlier Detection.

Abstract: Empirical data of RFC (response for a class), CBO (coupling between object classes), and WMC (weighted methods per class) software metrics, that can be used for estimation of software quality, deviate from normality. These metrics unveil multivariate skewness and kurtosis that do not conform to a multivariate Gaussian distribution. At the same time, well-known statistical methods that assume data normality may not be appropriate for the analysis of non-Gaussian data. To detect the outliers in the three-dimensional data of RFC, CBO, and WMC metrics and to estimate the confidence and prediction intervals of nonlinear regressions for these metrics, we need to use three-variate normalizing transformations. For statistical analysis of RFC, CBO, and WMC metrics, their normalization using the three-variate Box-Cox transformation was applied. Mardia's test for the transformed data after applying the multivariate Box-Cox transformation points that the transformed dataset is Gaussian. A technique for detecting outliers in multivariate non-Gaussian data based on the squared Mahalanobis distance for normalized data was applied to ensure the removal of outliers. Three nonlinear regression models for each of the RFC, CBO, and WMC metrics were constructed. The confidence and prediction intervals of nonlinear regressions for each of the RFC, CBO, and WMC metrics were built. Well-known statistical characteristics PRED(0.25) and MMRE for both the primary and the test datasets show that the model quality is satisfactory. The confidence and prediction intervals of nonlinear regressions for these metrics can be used for estimation of the quality of the object-oriented design of the software.

1 INTRODUCTION

Statistical analysis of multivariate data plays an important role in many areas, including empirical software engineering [1]. Empirical software engineering studies apply various methods of multivariate statistical analysis. Assuring the validity of such methods and corresponding results is challenging and critical [2]. As it is known [3], many methods of multivariate statistical analysis are based on the assumption that the data is normally distributed. Also, we know [4] if the data are not normally distributed, it is misleading to draw conclusions based on the normal distribution.

The above also applies to the well-known software metrics RFC (response for a class), CBO (coupling between object classes), and WMC (weighted methods per class). Although these metrics, along with three others (DIT - depth of inheritance tree, LCOM - lack of cohesion in methods, and NOC – a class's number of children), were proposed by Chidamber and Kemerer back in 1991 [5] for measuring the three non-implementation steps in Booch's definition of the object-oriented design (OOD), they are still used today to solve and other problems [6-15], including software quality [16-21]. In [21] the author proposed to apply the confidence and prediction intervals of nonlinear regressions for the RFC, CBO, and WMC metrics for evaluating the quality of software systems from the point of view of their OOD. In [21] the three-variate Box-Cox normalizing transformation was used to clean data from outliers and build the nonlinear regression models, the confidence and prediction intervals for the nonlinear regressions for the metrics RFC, CBO, and WMC since, firstly, according to the Mardia test, the three-dimensional data of these metrics are not normally distributed and, secondly, the residuals distribution of the corresponding linear regression models is not Gaussian. However, the above use of the three-variate Box-Cox transformation was based on the data sample from 51 open-source apps in Java.

In this paper, we have extended the results to a larger amount of data of metrics RFC, CBO, and WMC. As in [21], to detect outliers we apply the technique based on the squared Mahalanobis distance for the multi-dimensional normalized data and to build the nonlinear regression models, the confidence and prediction intervals for the nonlinear regressions for the metrics RFC, CBO, and WMC we use both univariate and multivariate normalizing transformations.

2 MATERIALS AND METHODS

Existing researches show that the data of software metrics can deviate from normality. Empirical values from the data set published in [21], consisting of RFC, CBO, and WMC metrics, were not normally distributed. These metrics unveil multivariate skewness and kurtosis that do not conform to a multivariate Gaussian distribution. At the same time, well-known statistical methods that assume data normality may not be appropriate for the analysis. Therefore, it is required to perform data normalization using normalizing transformations, as in [21]. Corresponding to [22], the bijective multivariate normalizing transformation will be used to convert a not Gauss-distributed random vector $\boldsymbol{P} = \{X_1, X_2, \dots, X_m\}^T$ into a Gauss-distributed random vector $\boldsymbol{T} = \{Z_1, Z_2, \dots, Z_m\}^T$:

$$\boldsymbol{T} = \Psi(\boldsymbol{P}), \quad (1)$$

where m is the number of metrics.

A transformation inverse to (1) is the following:

$$\boldsymbol{P} = \Psi^{-1}(\boldsymbol{T}). \quad (2)$$

In this research, we use multivariate Box-Cox transformation (BCT) (three-variate BCT in this case to transform values of each of RFC, CBO, and WMC metrics, taking into the correlation between the metrics):

$$Z_j = \begin{cases} \left(X_j^{\lambda_j} - 1\right)/\lambda_j, & \text{if } \lambda_j \neq 0; \\ ln(X_j), & \text{if } \lambda_j = 0. \end{cases} \quad (3)$$

There λ_j is a parameter of BCT and Z_j is the Gauss-distributed variable; $j = 1,2, \dots, m$. The estimates of these parameters are calculated by the method of maximum likelihood as in [3].

For the Box-Cox transformation the log-likelihood function is the following:

$$L(\boldsymbol{\lambda}) = -\frac{N}{2} ln\left(det(\boldsymbol{S}_N)\right) + (\boldsymbol{\lambda} - 1) \sum_{i=0}^{N} ln\,(X_i). \quad (4)$$

There $\boldsymbol{\lambda}$ is the vector of lambda values, $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \dots, \lambda_m\}$; N is the number of data rows; $\boldsymbol{S}_N$ is the sample covariance matrix:

$$\boldsymbol{S}_N = \frac{1}{N} \sum_{i=1}^{N} (\boldsymbol{T}_i - \overline{\boldsymbol{T}})\,(\boldsymbol{T}_i - \overline{\boldsymbol{T}})^T. \quad (5)$$

There $\overline{\boldsymbol{T}}$ is the sample mean vector, $\overline{\boldsymbol{T}} = \{\overline{Z}_1, \overline{Z}_2, \dots, \overline{Z}_m\}^T$; $\overline{T}_j = \frac{1}{N} \sum_{i=1}^{N} Z_{j_i}$; $j = 1,2, \dots, m$.

To ensure the removal of outliers, we apply a technique based on the squared Mahalanobis distance for normalized data, as described in [22]. For each multivariate data point the squared Mahalanobis distance d_i^2 can be calculated as

$$d_i^2 = (\boldsymbol{T}_i - \overline{\boldsymbol{T}})^T \cdot \boldsymbol{S}_N^{-1} \cdot (\boldsymbol{T}_i - \overline{\boldsymbol{T}}). \quad (6)$$

There $\boldsymbol{S}_N$ is the sample covariance matrix as defined in (5).

Regarding to [21], data points having d_i^2 that are greater than $3(N^2 - 1)F_{3,N-3,0.005}/N(N - 3)$ are treated as outliers. $F_{3,N-3,0.005}$ is the F-distribution quantile with 3 and $N - 3$ degrees of freedom and a significance level of 0.005.

All identified outliers must be removed if present.

After removing outliers it is possible to estimate the quality of data points. This requires constructing intervals of the prediction and of the confidence for non-linear regression models for each metric. As in [21], we will use the transformation inverse to (1) and regression analysis for the prediction interval construction:

$$\Psi_Y^{-1}\left(\hat{Z}_Y \pm t_{\frac{\alpha}{2},\nu} S_{Z_Y} \left\{1 + \frac{1}{N} + (\boldsymbol{z}_X^+)^T \boldsymbol{S}_Z^{-1} (\boldsymbol{z}_X^+)\right\}^{\frac{1}{2}}\right), \quad (7)$$

where Ψ_Y is the normalizing transformation component for the dependent variable Y; $\hat{Z}_Y$ is a result of prediction with the equation of linear regression, $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2$; $t_{\frac{\alpha}{2},\nu}$ is a quantile of the Student's t-distribution having ν degrees of freedom; $\nu = N - 3$; $\boldsymbol{z}_X^+$ is a vector consisting from $Z_{1_i} - \bar{Z}_1$, $Z_{2_i} - \bar{Z}_2$ for row i; $\bar{Z}_j = \frac{1}{N}\sum_{i=1}^{N} Z_{j_i}$, $j = 1, 2, \ldots, k$; $S_{Z_Y} = \frac{1}{\nu}\sum_{i=1}^{N}\left(Z_{Y_i} - \hat{Z}_{Y_i}\right)^2$.

In the (7) $\boldsymbol{S}_Z$ is the covariance matrix of the predictor variables, defined as (4):

$$\boldsymbol{S}_Z = \begin{pmatrix} S_{Z_1 Z_1} & S_{Z_1 Z_2} \\ S_{Z_1 Z_2} & S_{Z_2 Z_2} \end{pmatrix}, \quad (8)$$

where $S_{Z_q Z_r} = \sum_i^N \left[Z_{q_1} - \bar{Z}_q\right]\left[Z_{r_i} - \bar{Z}_r\right]$, $q, r = 1, 2$; $\bar{Z}_j = \frac{1}{N}\sum_{i=1}^{N} Z_{j_i}$, $j = 1, 2$.

For constructing intervals of the prediction for each of the involved metrics, we sequentially should treat one of the normalized metrics as the dependent variable of (7), and the remaining k metrics as independent variables of (7).

Regarding [21], for data points that are located inside the range between lower and higher values of the interval of the confidence (for each RFC, CBO, and WMC metrics) software quality is medium. For data points that are located inside the range from a lower value of the interval of the prediction to a higher value of the interval of the confidence for each metric, software quality is high. For other data points software quality is low.

3 ANALYSIS OF METRICS OF SOFTWARE DEVELOPED IN JAVA

An illustration of how we can use the listed methods of statistical analysis on the multivariate data when the data are not normally distributed will be provided. This illustration will be done by the estimation of the OOD quality of the software. Like in [21], we will use software-level RFC, CBO, and WMC metrics to construct three corresponding regression models, and intervals of the prediction and the confidence. In addition to the data described in [21] (53 rows), we will use 35 more data rows with empirical values of RFC, CBO, and WMC metrics for the software developed in Java. Additional rows are provided in Table 1.

These values were collected for the software developed in Java and stored in the public GitHub repositories. The collection of these values was performed by the CK framework on the class level, and then the values were converted to the software level by averaging the number of classes.

Table 1: OOD Metrics for the software developed in Java.

Repository	RFC	CBO	WMC
3D-TETRIS	9.906	4.547	21.516
Chemtris	5.234	3.455	4.977
Cubes	7.479	5.212	8.506
DestinationSol	12.923	8.344	12.34
finisterra	8.812	5.954	7.342
GDX-RPG	9.267	4.449	11.096
GdxGame	11.408	5.942	9.777
kickoff	8.068	5.227	8
Klooni1010	9.471	6.059	8.824
Koru	6.587	3.985	5.726
lightblocks	9.838	5.063	10.515
mario-game	9	5.022	9.109
martianrun	8.475	4.459	9.918
marvelous-bob	7.085	6.641	5.41
mini2Dx	10.156	5.039	14.939
Norii	15.193	6.62	17.422
Novix	10.06	3.595	5.888
OasisGame	9.373	6.152	8.088
odb-naturally-selected-2d	6.137	6.874	6.263
OverblownGame	7.94	6.508	10.377
Particle-Park	10.029	5.086	5.714

The first 3 rows from Table 1 will be merged with 53 rows from the dataset described in [21] (56 rows in total). 32 more rows from Table 1 will be used as a test dataset.

Checking the merged dataset with Mardia's test shows that this dataset is not Gaussian (multivariate skew $\beta_1 = 11.66$ and multivariate kurtosis $\beta_2 = 27.17$). Therefore, as expected, it is not possible to apply the method based on the squared Mahalanobis distance for unnormalized data to remove outliers and use linear regression analysis (the residual distribution of the corresponding linear regression models is not Gaussian).

Following [21], it is needed to use multivariate normalizing transformations to remove outliers from the data and construct non-linear regression models, and intervals of the prediction and the confidence. In this illustration, we will use multivariate Box-Cox transformation (3). Here we sequentially replace Z_j, X_j, λ_j with Z_{RFC}, X_{RFC}, λ_{RFC}, then with Z_{CBO}, X_{CBO}, λ_{CBO}, and finally with Z_{WMC}, X_{WMC}, λ_{WMC}.

The estimation of $\boldsymbol{\lambda}$ for (3) was done using the corresponding log-likelihood function (4). Estimates were calculated using the Apache Math library implementation of the BOBYQA Optimizer (Bound Optimization BY Quadratic Approximation).

Piece of a computer program in Java to get estimations of $\boldsymbol{\lambda}$ with BOBYQA:

```
public static double[]
estimateLambda(double[][] data) {
  int dimension = data[0].length;
  double[] initialGuess = new
double[dimension];
  Arrays.fill(initialGuess, 1.0);

  MultivariateFunction objective =
lambda -> computeLogLikelihood(data,
lambda);

  BOBYQAOptimizer optimizer = new
BOBYQAOptimizer(2 * dimension + 1);

  PointValuePair result =
optimizer.optimize(
    MaxEval.unlimited(),
    MaxIter.unlimited(),
    new ObjectiveFunction(objective),
    GoalType.MAXIMIZE,
    new InitialGuess(initialGuess),
    SimpleBounds.unbounded(dimension)
  );

  return result.getPoint();
}
```

Multivariate Box-Cox transformation parameter estimates for 56 rows are the following: $\hat{\lambda}_{\mathrm{RFC}} = 0.2562$, $\hat{\lambda}_{\mathrm{CBO}} = 0.7704$, $\hat{\lambda}_{\mathrm{WMC}} = -0.2198$.

Multivariate distribution after applying the multivariate Box-Cox transformation (3) with estimated parameters has a multivariate skew $\beta_1 = 1.53$ and multivariate kurtosis $\beta_2 = 15.45$. Specified value points that the transformed dataset is Gaussian. Therefore, it is possible to apply regression analysis on the transformed dataset, having previously checked for outliers.

Outliers check was done using (6) by calculating the squared Mahalanobis distance for each of the 3-dimensional rows. Sample means for this data are $\bar{Z}_{RFC} = 6.903$, $\bar{Z}_{CBO} = 2.086$, $\bar{Z}_{WMC} = 4.038$. The corresponding sample covariance matrix is the following:

$$\boldsymbol{S}_N^{-1} = \begin{pmatrix} 0.451 & 0.384 & -0.997 \\ 0.384 & 46.262 & -12.904 \\ -0.997 & -12.904 & 6.747 \end{pmatrix}.$$

The largest value of the squared Mahalanobis distance calculated with (6) is $d_{max}^2 = 12.08$. This value is smaller than a quantile of the F distribution with degrees of freedom 3 and 53 and a significance level of 0.005. It means that the transformed dataset does not contain significant deviations that can be treated as outliers.

By sequentially treating one of the normalized metrics as the dependent variable and the remaining 2 metrics as independent variables, we constructed three nonlinear regression models for each of the RFC, CBO, and WMC metrics based on 3-variate BCT:

$$Y = \left[\hat{\lambda}_Y\left(\hat{Z}_Y + \varepsilon\right)\right]^{1/\hat{\lambda}_Y}. \tag{9}$$

There $\hat{Z}_Y$ is a result of prediction with the equation of linear regression, $\hat{Z}_Y = \hat{b}_0 + \hat{b}_1 Z_1 + \hat{b}_2 Z_2$ that depends on predictors Z_1 and Z_2 for the Gaussian data transformed with 3-variate BCT; is ε is a Gaussian random variable, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$.

The estimates of the parameters for the nonlinear regression models (9) for each of the RFC, CBO, and WMC metrics are provided in Table 2. The distribution of ε of each linear regression for normalized metrics was Gaussian regarding the Chi-square test with a significance level of 0.05. Residual distributions and the density of the corresponding Gaussian distributions are shown in Figure 1.

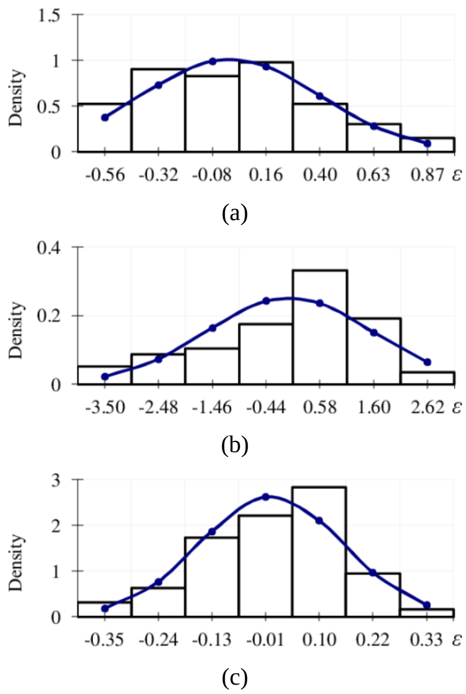

Figure 1: Residuals distributions of the linear regression models for the normalized values of RFC (a), CBO (b), and WMC (c) metrics.

Table 2: The estimates of the parameters for the nonlinear regression models.

No	Y	$\hat{b}_0$	$\hat{b}_1$	$\hat{b}_2$	σ_ε^2	MMRE	PRED
1	RFC	-0.972	0.148	1.913	0.389	0.16	0.839
2	CBO	-0.752	2.367	-0.913	1.555	0.241	0.679
3	WMC	1.017	0.279	-0.008	0.148	0.224	0.643

We used the popular characteristics, MMRE and PRED(0.25), to evaluate the predictive accuracy of nonlinear regression models (9) for each of the RFC, CBO, and WMC metrics. It is acceptable to have an MMRE of no more than 0.25 and a PRED(0.25) of no less than 0.75. The MMRE and PRED(0.25) values for the abovementioned models are shown in Table 2. These characteristics show that the model's quality is satisfactory.

To compute the intervals of the prediction and of the confidence for the nonlinear regression for the RFC metric the following values will be used in (7) and (8): $S_{Z_Y} = 0.157$, $\bar{Z}_1 = 6.903$, $\bar{Z}_2 = 2.086$, $\boldsymbol{S}_Z^{-1} = \begin{pmatrix} 0.005 & -0.027 \\ -0.027 & 0.385 \end{pmatrix}$.

To compute intervals of the prediction and the confidence for the nonlinear regression for the CBO metric the following values will be used in (7) and (8): $S_{Z_Y} = 2.510$, $\bar{Z}_1 = 4.038$, $\bar{Z}_2 = 2.086$, $\boldsymbol{S}_Z^{-1} = \begin{pmatrix} 0.078 & -0.214 \\ -0.214 & 0.820 \end{pmatrix}$.

To compute intervals of the prediction and the confidence for the nonlinear regression for the WMC metric the following values will be used in (7) and (8): $S_{Z_Y} = 0.0228$, $\bar{Z}_1$ =4.038, $\bar{Z}_2 = 6.903$, $\boldsymbol{S}_Z^{-1} = \begin{pmatrix} 0.056 & -0.016 \\ -0.016 & 0.007 \end{pmatrix}$.

To compute intervals of the prediction and of the confidence for the nonlinear regression for each of the RFC, CBO, and WMC metrics the following values will be used in (7) and (8): $N = 56$, $\nu = 53$, $t_{0.025,53}$=2.006.

For the test dataset consisting of 23 rows, the values of the MMRE and PRED(0.25) for the nonlinear regression models (9) for each of the RFC, CBO, and WMC metrics are provided in Table 3.

Table 3: Nonlinear regression models quality characteristics for the test dataset.

No	Y	MMRE	PRED
1	RFC	0.114	0.875
2	CBO	0.347	0.375
3	WMC	0.190	0.719

These characteristics for the test dataset also show that the model's quality is satisfactory.

4 DISCUSSION

For statistical analysis of RFC, CBO, and WMC metrics, we propose to apply their normalization using the three-variate Box-Cox transformation. This choice is due to the following. Firstly, the three-dimensional data of these metrics are not normally distributed and, secondly, the residuals distribution of the corresponding linear regression models is not Gaussian.

To detect the three-dimensional outliers in the data, we applied the appropriate technique [22] based on the multivariate normalizing transformations. The use of the three-variate Box-Cox transformation allows us to additionally take into account the correlation between RFC, CBO, and WMC metrics.

To build the confidence and prediction intervals for the nonlinear regressions for RFC, CBO, and WMC metrics for evaluating the quality of open-source apps in Java, we used a 0.05 significance level, as the appointed one usually, although this value may be discussed.

The statistical analysis of RFC, CBO, and WMC metrics based on the three-variate Box-Cox transformation demonstrates its capabilities. In the future, it is necessary to build corresponding mathematical models based on other data sets.

5 CONCLUSIONS

For statistical analysis of RFC, CBO, and WMC metrics, have proposed to apply their normalization using the three-variate Box-Cox transformation. To detect the outliers in the three-dimensional data of RFC, CBO, and WMC metrics and to estimate the confidence and prediction intervals of nonlinear regressions for these metrics, we need to use three-variate transformations. The constructed confidence and prediction intervals of nonlinear regressions for the above metrics may be applied to estimate the quality of open-source software in Java.

Moving forward, we plan to develop examples of statistical analysis of RFC, CBO, and WMC metrics that do not have limitations due to the programming language and the sample size.

ACKNOWLEDGMENTS

We are grateful to P. Fadeev, a bachelor's student at the Admiral Makarov National University of

Shipbuilding, for his assistance in collecting software metrics data, which allowed us to expand the dataset used in the data analysis.

REFERENCES

[1] D. Mendez, P. Avgeriou, M. Kalinowski, and N. bin Ali, Handbook on teaching empirical software engineering. Cham: Springer, 2024, doi: 10.1007/978-3-031-71769-7.

[2] J. Härtel and R. Lämmel, "Operationalizing validity of empirical software engineering studies," Empir. Softw. Eng., vol. 28, Nov. 2023, Art. no. 153, doi: 10.1007/s10664-023-10370-3.

[3] R. A. Johnson and D. W. Wichern, Applied Multivariate Statistical Analysis. New Jersey: Pearson Prentice Hall, 2007.

[4] J. Osborne, Best Practices in Data Cleaning: A Complete Guide to Everything You Need to Do Before and After Collecting Your Data. SAGE Publications, Inc., 2013, doi: 10.4135/9781452269948.

[5] S. R. Chidamber and C. F. Kemerer, "Towards a metrics suite for object oriented design," ACM SIGPLAN Notices, vol. 26, no. 11, pp. 197–211, 1991, doi: 10.1145/118014.117970.

[6] C. Haritha Madhav and K. S. Vipin Kumar, "A method for predicting software reliability using object oriented design metrics," in Proc. Int. Conf. on Intelligent Computing and Control Systems (ICCS), 2019, pp. 679–682, doi: 10.1109/ICCS45141.2019.9065541.

[7] E. A. Alomar, M. W. Mkaouer, A. Ouni, and M. Kessentini, "On the Impact of Refactoring on the Relationship between Quality Attributes and Design Metrics," in Proc. Int. Symposium on Empirical Software Engineering and Measurement (ESEM), Porto de Galinhas, Brazil, 2019, pp. 1-11, doi: 10.1109/ESEM.2019.8870177.

[8] M. Rizwan, A. Nadeem, and M. A. Sindhu, "Empirical Evaluation of Coupling Metrics in Software Fault Prediction," in Proc. 17th Int. Bhurban Conf. on Applied Sciences and Technology (IBCAST), 2020, pp. 434–440, doi: 10.1109/IBCAST47879.2020.9044489.

[9] A. Vescan, C. Serban, and G. C. Crisan, "Software defects rules discovery," in Proc. IEEE 14th Int. Conf. on Software Testing, Verification and Validation Workshops (ICSTW), 2021, pp. 101–109, doi: 10.1109/ICSTW52544.2021.00028.

[10] M. Begum, M. H. Shuvo, I. Ashraf, A. A. Mamun, J. Uddin, and M. A. Samad, "Software Defects Identification: Results Using Machine Learning and Explainable Artificial Intelligence Techniques," IEEE Access, vol. 11, pp. 132750–132765, 2023, doi: 10.1109/ACCESS.2023.3329051.

[11] M. Ziobrowski, M. Ochodek, J. Nawrocki, and B. Walter, "Towards Reliable Rule Mining about Code Smells: The McPython Approach," in Proc. 18th Conf. on Computer Science and Intelligence Systems (FedCSIS), 2023, pp. 65–66, doi: 10.15439/2023F2071.

[12] P. Ofem, B. Isong, and F. Lugayizi, "Metrics for Evaluating and Improving Transparency in Software Engineering: An Empirical Study and Improvement Model," SN Computer Science, vol. 5, Art. no. 8, 2024, doi: 10.1007/s42979-024-03471-3.

[13] A. S. Prykhodko and E. V. Malakhov, "Determining object-oriented design complexity due to the identification of classes of open-source web applications created using PHP frameworks," Radio Electronics, Computer Science, Control, 2024, no. 2 (69), pp 160–166, doi: 10.15588/1607-3274-2024-2-16.

[14] M.-A. Levasseur and M. Badri, "Prioritizing unit tests using object-oriented metrics, centrality measures, and machine learning algorithms," Innovations in Systems and Software Engineering, 2024, doi: 10.1007/s11334-024-00550-9.

[15] D. A. Rebro, S. Chren, and B. Rossi, "Source Code Metrics for Software Defects Prediction," in Proc. ACM Symposium on Applied Computing, 2023, pp. 1469–1472, doi: 10.1145/3555776.3577809.

[16] A.-J. Molnar, A. Neamţu, and S. Motogna, "Evaluation of Software Product Quality Metrics," in Proc. Communications in Computer and Information Science, vol. 1172, 2020, pp. 163-187, doi: 10.1007/978-3-030-40223-5_8.

[17] M. M. A. Dabdawb and B. Mahmood, "A Network of Object-Oriented Software Metrics Parameters," in Proc. 10th IEEE Int. Conf. on Communication, Networks and Satellite (Comnetsat), 2021, pp. 172–178, doi: 10.1109/COMNETSAT53002.2021.9530822.

[18] I. C. Nwandu, J. N. Odii, E. C. Nwokorie, and S. A. Okolie, "Evaluation of Software Quality in Test-driven Development: A Perspective of Measurement and Metrics," Int. Journal of Information Technology and Computer Science, vol. 14, no. 6, pp. 13–22, 2022, doi: 10.5815/ijitcs.2022.06.02.

[19] I. M. A. Wikantyasa, A. P. Kurniawan, and S. Rochimah, "C K Metric and Architecture Smells Relations: Towards Software Quality Assurance," in Proc. 14th Int. Conf. on Information and Communication Technology and System (ICTS), 2023, doi: 10.1109/ICTS58770.2023.10330874.

[20] S. Jin, Z. Li, B. Chen, B. Zhu, and Y. Xia, "Software Code Quality Measurement: Implications from Metric Distributions," in Proc. IEEE Int. Conf. on Software Quality, Reliability and Security (QRS), 2023, pp. 488–496, doi: 10.1109/QRS60937.2023.00054.

[21] S. Prykhodko, "Evaluating Quality of Software Systems by the Confidence and Prediction Intervals of Regressions for RFC, CBO and WMC Metrics," WSEAS Transactions on Systems, vol. 23, pp. 322–330, 2024, doi: 10.37394/23202.2024.23.36.

[22] S. Prykhodko, N. Prykhodko, L. Makarova, and A. Pukhalevych, "Application of the squared Mahalanobis distance for detecting outliers in multivariate non-Gaussian data," in Proc. 14th Int. Conf. on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), 2018, pp. 962–965, doi: 10.1109/TCSET.2018.8336353.