ISSN 1870-4069

Artificial Intelligence and Component-Based Software Engineering: A Systematic Mapping Study

Bruno A. López-Luján, Ángel J. Sánchez-García, Karen Cortés-Verdín

Universidad Veracruzana,
Facultad de Estadística e Informática,
Mexico

{angesanchez, kcortes}@uv.mx, bruno1529@live.com.mx

Abstract. Industry 4.0 has changed the way businesses and organizations operate. Faced with this change, organizations must identify the technologies that best meet their needs in order to invest in them, especially in the way of developing software products that meet the needs that this revolution involves. Component-Based Software Engineering (CBSE) seeks reusing to define, implement, and compose weakly coupled software components in systems. This research proposes to carry out a Systematic Mapping Study (SMS), based on the Kitchenham and Charters guide, to associate Artificial Intelligence (AI) approaches to CBSE problems. In this SMS, 38 primary studies were selected. From these studies, it was observed that Artificial Intelligence has supported to CBSE through 16 algorithms and 17 AI techniques, highlighting optimization techniques such as Evolutionary Algorithms. On the other hand, 28 problems were found within 14 CBSE activities where the activity most addressed by AI has been Software Reliability.

Keywords: Artificial intelligence, component-based software engineering, systematic mapping study, optimization.

1 Introduction

Industry 4.0 is identified by the emergence of new technologies such as Robotics, Analytics, Artificial Intelligence (AI), Nanotechnology and the Internet of Things (IoT), among others. However, all these areas converge in the need for software development that computes and manages the data that is generated every day.

Recently, Software development has seen in Component-Based Software Engineering (CBSE) an opportunity to reduce software development times and therefore release and delivery times. CBSE is “an approach that proposes the development and evolution of a software system through the selection and composition of components" [1].

In addition, CBSE is known to be a specialized way of creating software from existing components, allowing the reuse of previously used components, where these components are selected for use. On the other hand, Artificial Intelligence (AI) is a very

broad field that began shortly after the Second World War. Throughout history, AI has followed four approaches: systems that think like humans, systems that think rationally, systems that act like humans, and systems that act rationally [2].

The Software Engineering and Artificial Intelligence disciplines have collaborated with each other in various areas, such as Requirements Classification [3], Requirements Prioritization [4], Component-Based Software Engineering [5], among others. There are Artificial Intelligence techniques that are used to solve some of the problems currently present in the different activities of Component-Based Software Engineering.

The applications of AI techniques in CBSE have shown successful results [6, 7, 8] or at least, raise proposals for future research [9, 10]. However, the literature indicates that Component-Based Software Engineering has problems that, although there are proposals to solve them, have not been fully resolved. Performing a manual search, no research papers have been found that succinctly cover all the applications of AI techniques for solving the problems that exist within Component-Based Software Engineering.

This is important since researchers currently have difficulty obtaining information from the AI techniques used in CBSE, so this work will provide a reliable source of information for them. The Systematic Mapping Study (SMS) will allow knowing the current state of Artificial Intelligence applications in Component-Based Software Engineering. In addition, it will benefit researchers, to strengthen the domain of the subject with the applications, the result of said applications, problems addressed, and proposals and solutions found.

This in order to facilitate the creation of research works in this area and the development of innovative contributions. This paper is organized as follows: Section 2 presents the background and related work. In Section 3, the method followed to carry out this SMS is described. Section 4 presents the final results. Finally, Section 5 draws conclusions and defines future work.

2 Background and Related Work

In the manual search of related works, and according to the authors' knowledge, only a paper was found as a systematic mapping study related to the subject. In 2018, Diwaker et al. [11] carried out a research where they made a list of all the applications of techniques and algorithms of the Soft Computing category belonging to Artificial Intelligence on different problems (mainly the prediction of software reliability) found within the Software Reliability activity pertaining to Component-Based Software Engineering.

The list of the soft computing techniques found were “Genetic Algorithm (GA), Neural-Network (NN), Fuzzy Logic, Support Vector Machine (SVM), Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Artificial Bee Colony (ABC)” [11] (only soft computing).

The research covered the needs, the area of opportunity that exists, as well as the difficulties and problems it presents, and the problems that have already been addressed in Software Reliability. This text describes the techniques and algorithms that have been used in Software Reliability, approached from Soft Computing. Finally, the importance of activities within the CBSE was exposed.

Table 1. Keywords.

Concept	Keywords
Component-Based Software Engineering	Component-Based Software Engineering, Component Based Software Engineering, Component-Based Software, Component Based Software
Artificial Intelligence	Artificial Intelligence, methodology, algorithm, technique

Table 1. Inclusion criteria.

IC	Description
IC1	The paper published from 2011 to May 2021 will be included
IC2	The paper that is written in English will be included.
IC3	The paper that in the title or abstract includes at least two key search concepts will be included.
IC4	The article that after reading the abstract gives the impression that it solves the research questions will be included.

3 Research Method

To carry out this research work, the guidelines for Systematic Literature Reviews (SLR) in Software Engineering, which were proposed by Kitchenham & Charters [12], were used as a guide. Although this guide is focused on SLR, it was decided to carry out a Systematic Mapping Study (SMS) to relate Artificial Intelligence techniques with CBSE problems as shown in the research questions. The method is presented in two stages: planning and execution.

3.1 Planning

Research Questions. The research questions (RQs) relating to this SMS are:

- **RQ1:** What are the Artificial Intelligence techniques that have been reported in the research of the CBSE area?
- **RQ2:** What have been the results in the applications of these techniques?

The motivation of RQ1 is to present the current state of AI algorithms and techniques applications at the CBSE. On the part of the RQ2, the motivation is to present the results of the applications of AI techniques and algorithms on the activities and problems of the CBSE to identify areas of opportunity.

Search strategy. An automated search was performed, searching for the primary studies among the resulting papers that demonstrated the selected information sources. All of this was done from a defined search string.

Keywords. Once the research questions were defined, the keywords were defined as shown in Table 1. The words "Artificial intelligence" and "Component-based software

Table 3. Exclusion criteria.

EC	Description
EC1	Papers that are not available will be excluded.
EC2	Papers that are only available in the form of slides, posters, books and technical reports will be excluded.
EC3	Papers with duplicate research will be excluded.

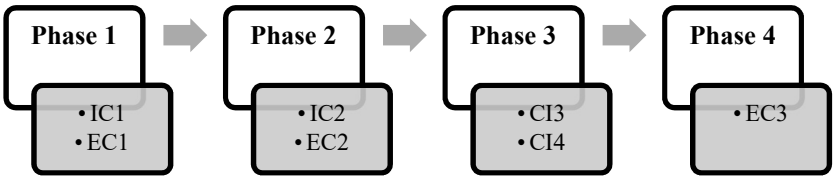

Fig. 1. Primary studies selection process.

engineering" were located as the most important terms for this research. Similar or related terms were added for each search keyword. The terms AI and CBSE were not included since not all authors use these abbreviations.

Search string. The keywords allowed the creation of the search string that was used in this investigation. After trying different chains of remaining as follows:

("Component-Based Software Engineering" OR "Component-Based Software" OR "Component Based Software" OR "Component Based Software Engineering") AND ("Artificial Intelligence" OR "Methodology" OR "Algorithm" OR "Technique")

Information sources. The information sources used for the search and selection of primary studies were four databases, being digital libraries: IEEE Xplore, ACM Digital Library, ScienceDirect and SpringerLink. These information sources were selected because they are repositories of articles from the computer science area and related disciplines.

Primary study selection criteria. Table 2 and Table 3 show the inclusion and exclusion criteria used to select the primary studies.

Primary study selection procedure. The phases that were applied to the resulting papers in the automated search for the selection of primary studies are found in Fig. 1.

Quality assessment. The quality evaluation was considered feasible to avoid threats to the quality of the information obtained as results of this research. For this point, the quality evaluation criteria were defined to have a quantifiable and objective result shown in Table 4.

These criteria are made up of eight questions applicable to each paper selected as the primary study, where 1 point will be given if it meets the criterion quality assessment or 0 if not compliant. The decision was made to select the papers that have a minimum of five points for use in this research. In this way we affirm that the articles used in this research are reliable and of quality.

Table 2. Quality evaluation criteria.

ID	Question
1	Are the objectives of the paper described?
2	Does the study mention papers or proposals already published?
3	In the related work section, does the study mention at least 3 papers that have been published up to five years before publication?
4	Does the study use public datasets?
5	Is the methodology used described in the study?
6	Was the research process adequately documented in the study?
7	Do the results prove the achievement of the objectives?
8	Are the conclusions of the study clear?

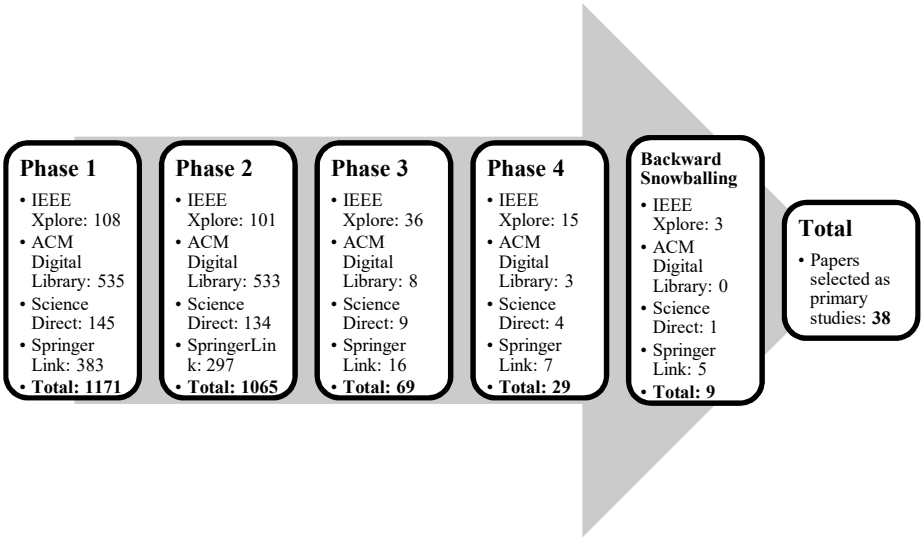

Fig. 2. Primary studies selection phases.

Data extraction. The template shown in Table 5 shows the data extracted from each primary study. Two types of data were extracted from each primary study and deemed necessary to answer the research questions.

The first seven data correspond to the details of the publication and the last two data are directly linked to RQ1 and RQ2, these data can answer both research questions.

3.2 Execution

The search string applied in the information sources yielded a total of 6,452 papers. As shown in Fig. 2, after applying all the phases of the selection process for primary studies, a result of 29 papers was obtained. Once the papers resulting from the selection procedure of primary studies shown in Fig. 1 were obtained, *Backward Snowballing* was applied to each of them, giving a result of 9 papers. Finally, a total of 38 papers

Table 5. Data extraction template.

Kind of data	Data extracted
Publication details	Title
	Authors
	Year
	Source
	Publication type
	Reference
	Abstract
Context	AI technique used
	CBSE activity and problematic addressed

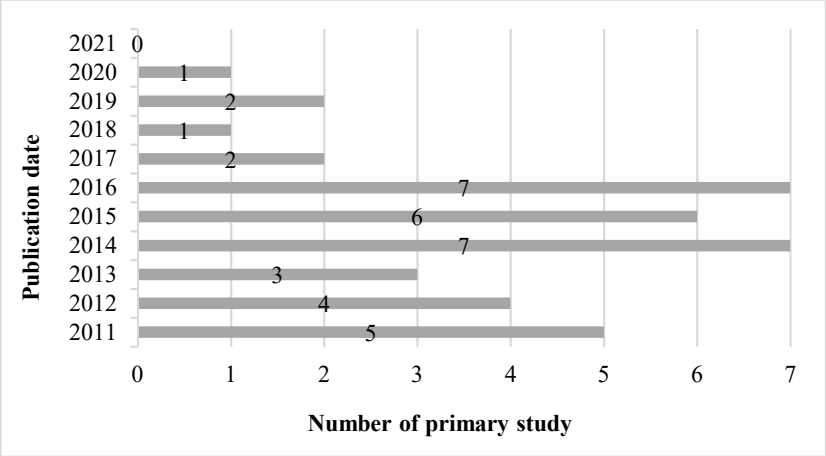

Fig. 3. Number of papers per year publication.

selected as primary studies were obtained. The detail of these 38 primary studies is presented in Appendix A[1].

4 Results

Of a total of 6,452 papers that emerged from the application of the search string in the four information sources, only 38 papers were selected as primary studies. In Fig. 3, a trend can be seen by researchers in the publication of papers where CBSE is addressed using AI techniques in the years 2014 to 2016.

Only in these three years were published 20 of the 38. To answer the research questions, a Thematic Synthesis Process was followed. Thirty-eight papers were selected as primary studies.

[1] Appendix A. 38 primary studies:
https://drive.google.com/file/d/1NqaIAs5cc1fhHPov3Zdx4M3fhxYjh6Sm/view?usp=sharing

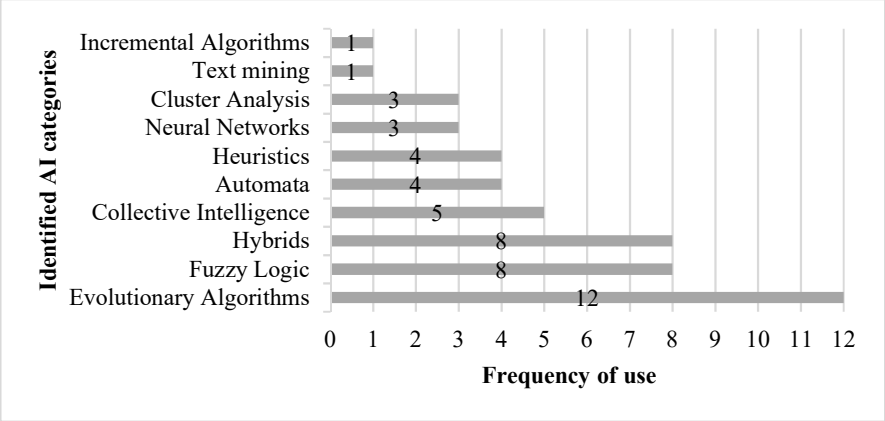

Fig. 4. Classification of AI algorithms and techniques.

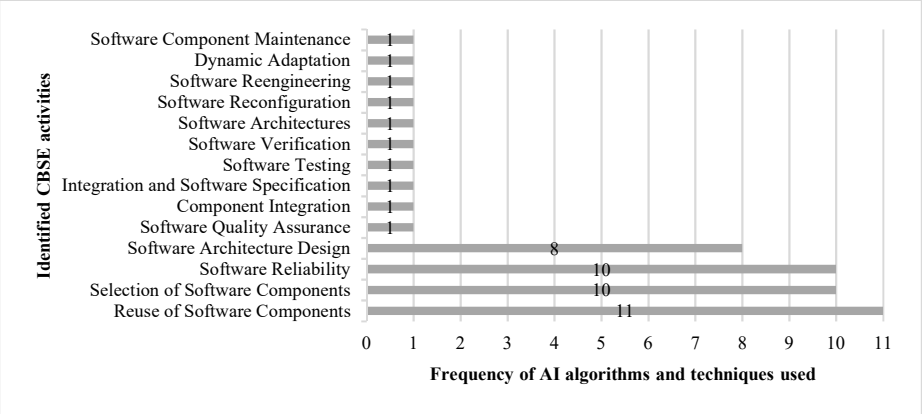

Fig. 5. CBSE activities and boarding frequency.

In these papers, 16 algorithms and 17 AI techniques were found to address different problems within the activities of the CBSE. The contributions of the AI on the CBSE were classified by categories.

To achieve this, techniques and algorithms that belonged to the same family were grouped as *Fuzzy Inference System* (FIS) and *Fuzzy Formal Concept Analysis* (FCA).

Both techniques were classified in *Fuzzy Logic*. To cite another example, *Evolutionary Algorithms* such as *Strength Pareto Evolutionary Algorithm* (SPEA2) and *S-Metric Selection Evolutionary Multiobjective Algorithm* (SMS-EMOA), as well as the rest of the evolutionary algorithms were classified in the same category. This classification can be seen in Fig. 4. Furthermore, the frequency of use of each of these categories on the CBSE can be seen. Regarding the problems and activities of the CBSE.

Within the primary studies 14 CBSE activities were found that were addressed using AI algorithms and techniques. Fig. 5 shows the activities found and the frequency of the AI approach. In Fig. 6, the problems within the identified activities are exposed. A trend is shown by researchers in the Reuse of Software Components approach, being

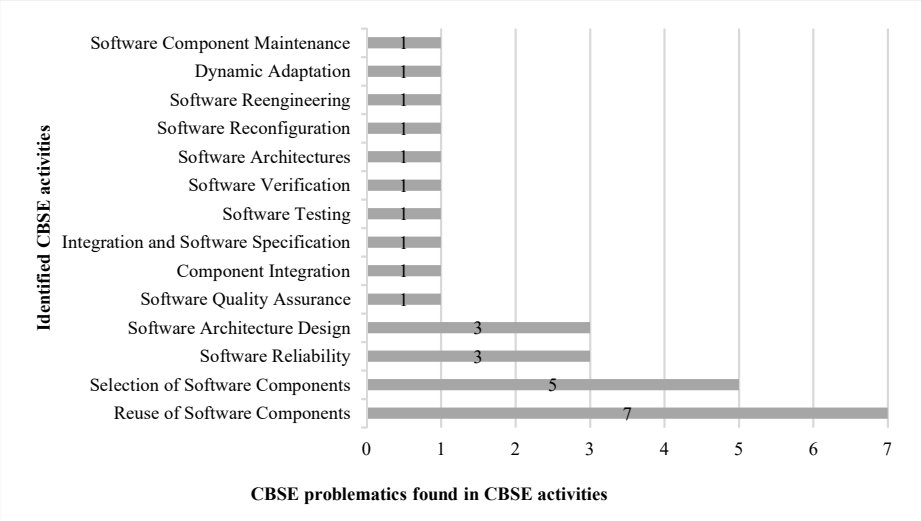

Fig. 6. Problematics found in CBSE activities.

the activity, most approached with proposals aimed at solving 7 problems. In response to RQ1, the use of 16 algorithms and 17 AI techniques on different issues within the activities of the CBSE was found.

We can expose the use of Genetic Algorithms (GA) which reported four approaches to solve different problems of the CBSE. Adaptive Neuro-Fuzzy Inference System (ANFIS), like GA reported four approximations. Finally, Fuzzy Logic (FL) with three proposals, together with GA and ANFIS add up to just over 33% of the proposals to solve problems or optimize CBSE activities.

This analysis demonstrates the variety of AI algorithms and techniques that have been reported in the current state of applications in the CBSE area, since the two techniques and the aforementioned algorithm were the most used by researchers in this research work.

As shown in Fig. 7. Evolutionary Algorithms (EA) were the most common algorithms for the approach to software reliability and the selection of the software component. However, in this category it is also proposed to address Software Testing and Software Architectures. In this way, Algorithms of evolution proves to be the category with flexibility in the contributions of AI on CBSE.

On the other hand, Text Mining and Incremental Algorithm only reported a contribution to address the activities of the CBSE, these being Integration and Specification of Software and Selection of Software Components respectively. Based on the results obtained from the Systematic Mapping Study and to respond to the RQ2, 38 studies are reported that address the CBSE through AI to address different activities and problems of this.

Each of the studies clearly describes its objectives and concisely explains the approach to a CBSE activity through a solution proposal, as well as other previously published studies to take them as a reference or compare them with their solution proposals. However, not all the selected primary studies mention updated articles

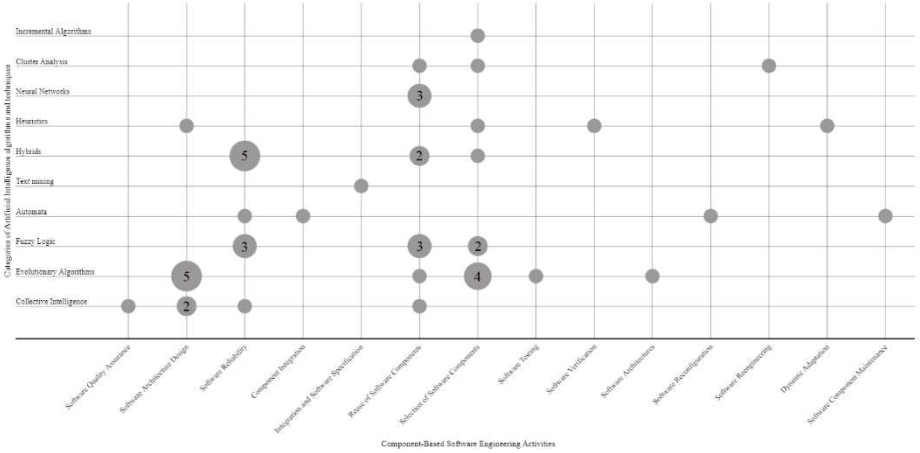

Fig. 7. Frequency of approach of AI on CBSE.

(taking as updated papers those papers published up to five years before the publication of the primary study) for comparison.

Selected primary studies use specific datasets, but not all datasets used are public or freely accessible. On the other hand, the selected primary studies describe the proposed methodology textually, graphically and concisely for a greater understanding of the reader. They also adequately address the proposal research process in each of them.

5 Conclusions and Future Work

A Systematic Mapping Study (SMS) was carried out to answer two research questions, applying the method proposed in [12] in four sources of information, where 38 articles selected as primary studies reported contributions on CBSE using AI techniques and algorithms. Once the results of this research work have been analyzed, classified and compared, it can be noted that there is a great tendency for researchers in CBSE to use Evolutionary Algorithms, Fuzzy Logic techniques, as well as algorithms and heuristic techniques to address problems pertaining to the different activities of Component-Based Software Engineering.

In addition, this research shows that the CBSE activities most addressed by Artificial Intelligence are: Component Reuse, Component Selection, and Software Reliability. The results obtained indicate that Artificial Intelligence addressed Component-Based Software Engineering 38 times through 16 algorithms and 17 AI techniques, solving 28 problems found within 14 CBSE activities.

In this work, it was shown that Artificial Intelligence and Software Engineering, specifically the area of Component-Based Software Engineering. AI supports several activities in the CBSE area, such as the choice of the best component, reliability estimation models, text mining, among others, which benefit in better reuse, less expensive and more accurate.

So, there is a lot of research to be done in this field in collaboration between areas, to ensure faster and more agile software development, as well as high quality, which is

fundamental in Industry 4.0. As future work, it is proposed to explore different optimization approaches to compare the results with those provided in the primary studies. This is because although they are addressed problems, they are not solved problems.

References.

1. Vale, T., Crnkovic, I., Santana de Almeida, E., da Mota Silveira Neto, P. A, Cavalcanti, Y. C., Romero de Lemos Meira, S.: Twenty-eight years of component-based software engineering. Journal of Systems and Software, vol. 111, pp. 128–148 (2016)
2. Russell, S. J., Norvig, P., Davis, E.: Artificial intelligence: A modern approach. Upper Saddle River, NJ: Prentice Hall (2003)
3. Hey, T., Keim, J., Koziolek, A., Tichy, W. F.: NoRBERT: Transfer learning for requirements classification. In: IEEE 28th International Requirements Engineering Conference (RE), pp. 169–179 (2020)
4. Kifetew, F., Munante, D., Perini, A., Susi, A., Siena, A., Busetta, P.: Dmgame: A gamified collaborative requirements prioritization tool. In: IEEE 25th International Requirements Engineering Conference (RE), pp. 468–469 (2017) doi: 10.1109/R E.2017.46
5. Xie, T.: The synergy of human and artificial intelligence in software engineering. In: 2nd International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering (RAISE), pp. 4–6 (2013)
6. Su-Wei, G.: Software component retrieval method based on PSO-RBF neural network. In: 2nd International Conference on Computer Engineering and Technology, vol. 7, pp. V7–339 (2010)
7. Khatri, S. K., Kaur, G., Johri, P.: Multi-level selection of reusable software components. In: Proceeding of 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 67–71 (2016)
8. Khode, S. G., Bhatia, R.: Improving retrieval effectiveness using ant colony optimization. In: Proceeding of International Conference on Advances in Computing, Control, and Telecommunication Technologies, pp. 737–741 (2009)
9. Cortellessa, V., Potena, P.: How can optimization models support the maintenance of component-based software? In: Proceeding of 1st International Symposium on Search Based Software Engineering, pp. 97–100 (2009)
10. Vodithala, S., Pabboju, S.: A dynamic approach for retrieval of software components using genetic algorithm. In: Proceeding of 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), pp. 406–410 (2015)
11. Diwaker, C., Tomar, P., Poonia, R. C., Singh, V.: Prediction of software reliability using bio inspired soft computing techniques. Journal of medical systems, vol. 42, no. 5, pp. 1–16 (2018)
12. Kitchenham, B., Charters, S.: Guidelines for performing systematic literature reviews in software engineering. Durham, UK: University of Durham (2007)