Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.017

IJCSMC, Vol. 6, Issue. 12, December 2017, pg.20 – 33

Approaches in Regression Testing Based on Test Case Prioritization: A Systematic Literature Survey

Kavya B.N, Manoj B.N
Hobvision Technologies, India
BCA, Shesadripuram Main College, India
kavyanarayan312@gmail.com, manojnarayan0704@gmail.com

Abstract— Soft-ware quality can be assured by going through soft-ware testing process. However, soft-ware testing phase is an expensive process as it consumes a longer time. By scheduling test cases execution order through a prioritization approach, soft-ware testing efficiency can be improved especially during regression testing. It is a notable step to be taken in constructing important soft-ware testing environment so that a system's commercial value can increase. The main idea of this review is to examine and classify the current test case prioritization approaches based on the articulated research questions. Test case prioritization has already been considerably discussed in the soft-ware testing do- main. However, it is commonly learned that there are quite a number of existing prioritization techniques that can still be improved especially in data used and execution process for each approach.

Keywords— "Testing", "Test cases", "Prioritization", "SLS", "TCS".

I. INTRODUCTION

Soft-ware engineering is not just programming and soft-ware development. Soft-ware engineering itself is an implementation of engineering procedures in the development of any soft-ware in a systematic way [1]. Within a soft-ware development process, soft-ware testing consumes a longer time in execution and can be the most expensive phase [2]. Soft-ware testing itself is normally, repetitively, carried out even when there are time constraint and fixed resources. Soft-ware engineering groups are regularly compelled to end their testing activities because of financial and time necessities, which will trigger some difficulties such as problems with the soft-ware quality and client agreement. However, the application of test case prioritization (tcp) appears to enhance test viability in soft-ware testing activity [3]. Regression testing is an activity to confirm that progress-ions do not harm the previously functioning soft-ware [4, 5]. As the soft-ware evolves, a soft-ware test suite has the tendency to increase in size which frequently makes it expensive to execute. Research shows regression testing is an expensive process which may require more than 33% of the cumulative expenses of the soft-ware [6]. In the work of yoo and harman [7], various regression test approaches were examined to supplement the importance of the accumulated test suite in regression testing. Those studies were then classified into three domains; minimization, selection, and prioritization. Test suite minimization (tsm) approaches intend to distinguish repetitive experiments and to eliminate test cases from a test suite execution with a specific goal such as to decrease the number of tests to run [8]. Minimization is sometimes called 'test suite reduction', meaning the elimination of test cases is permanent. Test case selection (tcs) approach also aims to decrease the number of test cases to be executed; however, the main idea of selection approach is that it is intended to be

modification-aware [9]. Tcs tries to recognize the test cases which would be important to the latest changes on soft-ware.

Lastly, test case prioritization (TCP) aims to order a set of test cases to achieve an early optimization based on preferred properties [3, 10]. It gives an approach the ability to execute highly significant test cases first according to some measure, and produce the desired outcome, such as revealing faults earlier and providing feedback to the testers. It also helps to find the ideal permutation of a series of test cases and could be executed accordingly [7]. Test case minimization reduces the test case amount in a set of test suite continuously while selection technique performs a temporary selection of several test cases which related are to modification awareness. From selective selection, important test cases might be missed out from the test suite. These test cases could possibly contain an important priority that needs to be executed to reveal certain faults. In test case prioritization, every single test case including new test cases that are added into present test suite execution will be considered in prioritization. This is crucial as new test cases will be executed to test a modified part of the soft-ware, hence, any abnormalities in the functional output could easily be observed. Despite the fact that there are numerous TCP approaches in the literature, there are no latest progressive literature reviews which illustrate recent TCP importance in soft-ware testing re- search. Therefore, this review attempts to perform a systematic literature survey (SLS) on the latest TCP approaches as proposed by Kitchenham [11]. SLS is a specialized, uncompromising, study of research evidence [12]. The point of an SLS is not to simply summarize all current proofs based on research questions; it is also expected to bolster the improvement of evidence-based research recommendations for researchers. This systematic literature survey is structured as follows. Section 2 considers the previous studies related to TCP approaches. Section 3 describes the strategy embraced to direct this SLS. Next, result and discussion based on the research questions were discussed in Section 4. Research findings were then elaborated in Section 5. In Section 6, the threat of validity to this SLS was discussed. Finally, Section 7 presents conclusion with regard to this systematic literature survey.

II. BACKGROUND STUDIES OF TEST CASE PRIORITIZATION

This section will discuss the previous studies that are related to TCP in regression testing. There were a few systematic reviews originated under the regression test case prioritization techniques domain. From the literature gathered, the authors collated one SLS, one mapping study, and three survey studies that are related to regression testing and TCP. The only SLS, work by Singh [13] , offered a systematic review in regression test case prioritization study covering the time period from 1997 to 2011. In their work, from 65 studies, 49 were identified to initiate a different approach, two on augmentation of prior studies, and 14 on analysing back earlier testified study results. The SLS also analysed and identified about eight broad prioritization approaches. These approaches include; genetic-based, coverage, requirement, modification, history, fault, composite, and others which include several approaches. The SLS concludes that even as there were different kinds of approaches, the main objective of TCP in regression testing remains the same, which is to increase fault detection.

On the other hand, in the work of [14], the authors presented a systematic mapping in test case prioritization with a specific focus on TCP studies. It covered the time period between 2001 and 2011. A majority of the reviews recorded in their work were about the approval of different looks into TCP and solution recommendations. The results are the same as reported in the previous SLS with a similar inadequacy. The authors manage to identify 16 studies out of 120, which correlated with the strength and weakness of some prioritization techniques. To locate the finest prioritization method, additional reviews on the analysis of prioritization systems are needed to be done as different approaches are constantly being produced. In addition to this SLS and one mapping study, there is one significant survey in regression testing domain and two minor surveys. The first significant survey is by Yoo and Harman [7] , focusing on regression testing area which includes test suite minimization (TSM), test case selection (TCS), and test case prioritization (TCP). For TCP, they classified the current state-of-the-art approaches into several categories. Those approaches were classified based on requirement, model, coverage covered, historical data, probabilistic calculation, cost awareness, and others which include several minor approaches. The authors analysed 159 studies covering the time period between 1977 and 2007. The authors also mentioned that experimental designs and evaluations process were still not harmonized but it is getting the attention of some researchers lately. In the first minor survey [15], this only analysed 19 studies and only focused on coverage-based technique and cost-effective technique. The authors of the survey suggested that having a new technique applied during the early stages of soft-ware development life cycle may significantly reduce development cost. The next minor survey [16], identified 90 papers from IEEE publication which summarized that there is a need to have an understanding of cost components and the advantages of having different parameters that could be taken into account. To summarize the background studies in previous related works, Table 3 shows the summary of findings in related studies in comparison to this SLS paper. From Table 3, it can be seen that only one study provides an empirical evidence for some TCP approaches, while other studies only provide an overview of some approaches. Therefore, there is an incomplete detailed overview such as the reasons behind the trends regarding some TCP approaches, which

need to be covered. It also can be noticed that most of the works only summarize the number of usage for each evaluation metric, but did not include in-depth discussions. Work by Singh [13] covered evaluation metrics but the gaps regarding usage of artefacts on specific TCP approach were not well discussed. In short, there are several uncovered findings that can be added to the current SLS work.

III. RESEARCH METHOD USED

With a specific end goal, a structured method to perform this SLS, as shown in Fig. 1, was implemented in order to examine the studies that are related to TCP. The systematic and structured method was inspired by Kitchenham [11, 12] and Achimugu [17]. Referring to Fig. 1, there are five main phases within the review protocol, itemized as follows. Research Question, Selection of Repositories, Search Strategy, Study Selection, and Data Synthesis and Extraction. In the first phase, four main research questions were generated to answer the main aim of this paper review. Next, selection of relevant repositories was performed. This is followed by employing a search strategy comprising specifying search strings and search process, which were planned based on the articulated research questions. The output of the search stage was then moved into the study selection phase. In this phase, the outcome of the search process underwent inclusion and exclusion criteria scrutiny to extract relevant studies. Quality assessments were then applied to evaluate the scrutinized studies further. Finally, the last phase dealt with data synthesis and extraction of primary studies utilized for this SLS.

A. Research questions and their motivations

This SLS aims to comprehend and review recent experimental evidence with respect to the most recent prioritization approaches in TCP area for further investigation; keeping in mind the end goal is to improvise the ability of present approaches. At the same time, the authors wish to review the empirical evaluations used in each reviewed approach. To accomplish this goal, four research questions with respective motivations were articulated as presented in Table 1. All these research questions, that frame the reason for under-taking this research are relatively connected and concurrently explored. These research questions are used to answer the extra findings that will be covered in this SLS. To be clear, Table 2 maps each research question to its respective extra finding and the finding's significance.

From Table 2, each RQ answers some uncovered findings from previous works, except for RQ 3. For each RQ, the questions are not only designed to answer the uncovered findings, but they are used to cover some extra findings that can be added to this SLS study. The significance of the findings for each RQ has also been detailed out as guidance to achieve the goal of this SLS study.

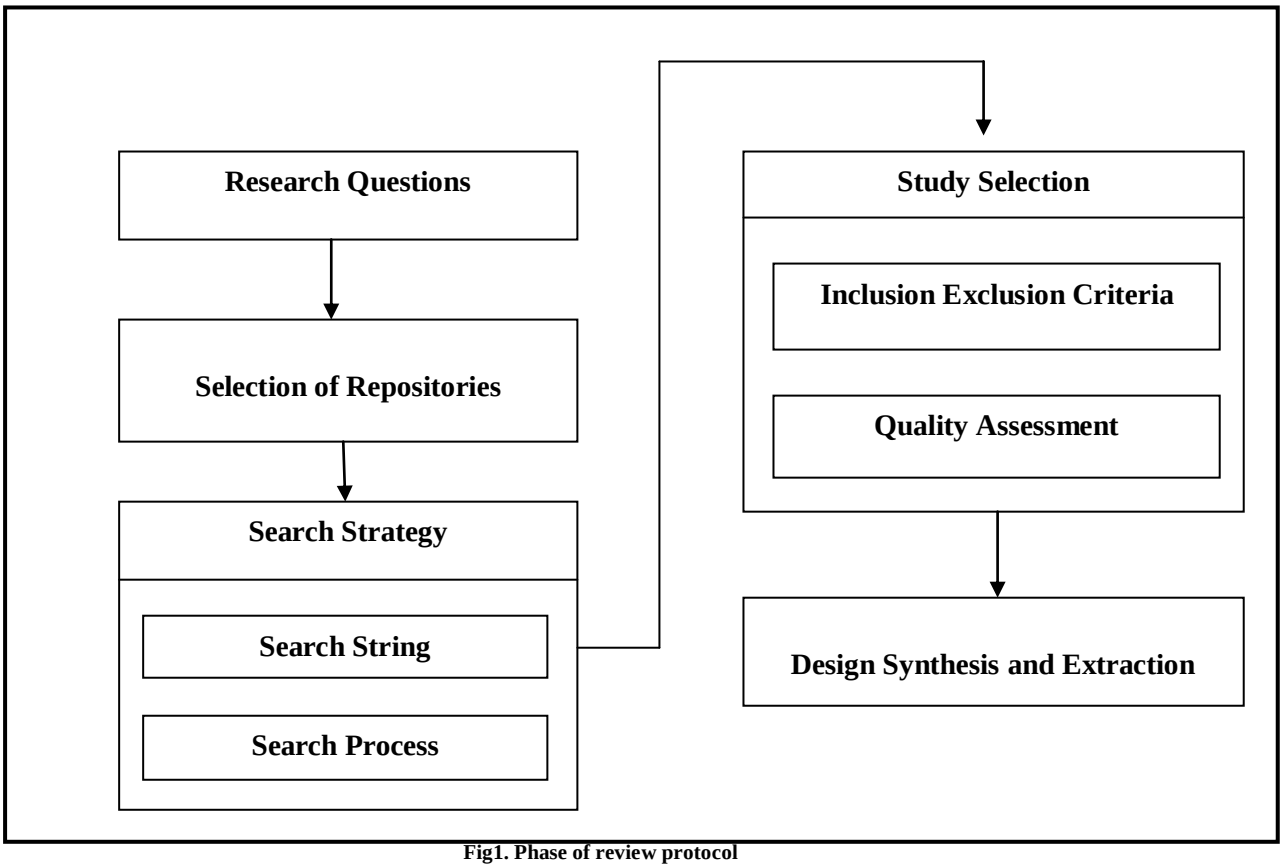


Fig1. Phase of review protocol

Research Questions	**RQ statement**	**Motivation**
RQ 1	What are the taxonomies of test case prioritization in regression testing?	These research questions focus on characterizing the current domain of test case prioritization. The reason is to know the development of TCP in regression testing throughout the past years.
RQ 1.1	What is the research trend of TCP techniques in regression testing?	
RQ 1.2	What are the distributions of approaches in TCP techniques?	The knowledge of differences in approaches is necessary to give a glimpse on how each prioritization approach functions, while the strength and weakness serve as the basis for improvement.
RQ 2	What are the differences in terms of approaches for each TCP technique?	
RQ 2.1	What are the descriptions, strength, and limitations of existing prioritization approaches?	
RQ 2.2	How were these approaches applied and how did they affect TCP results?	This research question can also help to illustrate the basic process of TCP execution in all different approaches.
RQ 3	What are the processes involved in TCP in regression testing?	
RQ 4	What are the evaluation metrics and suitable types of artefacts involved in TCP with the reasons for their creation?	This research question helps other researchers to choose which evaluation metric is suitable for their controlled experiment or case study.

Table1. Research questions and their motivations

Research questions	**Uncovered findings answered**	**Extra findings**	**Significance of the findings**
RQ 1	Reasons behind the trends of each TCP approaches	Distributions of approaches in TCP techniques and its logic	To provide insights in the development of TCP in regression testing within the recent trends.
RQ 2	A detailed overview of other recent popular TCP approaches Empirical evidence for other recent TCP approaches	The strength and limitations of existing prioritization approaches	To provide a glimpse of how each prioritization approach functions, and serve information for improvement
RQ 3	.....	Processes involved in TCP	To illustrate the basic process of TCP execution for improvement Information of available evaluation metrics and artefacts with the reasons of creation and relation to a specific TCP approach
RQ 4	An uncovered evaluation metric available with its reasons of creations Gaps regarding the usage of artefacts on TCP approach	Metric most likely used for specific TCP approach	

Table2. Mapping of research questions to uncovered finding with its significance

B. Study strategy

A study strategy is crucial in every SLS to guarantee the broadness of the selected studies. The value of the SLS is generally realized according to the selected primary studies. Strategy for this survey depended on these three stages:

a) Literature repository selection
b) Search string identification
c) Study selection process

C. Literature repository selection

The authors initiated this selection process by entering 'Test Case Prioritization' as search strings with the exact phrase on Google Scholar database. This database returns 2760 of studies available. From the search result, the authors identified several popular repositories in TCP research area and decided to gather the primary studies originated from recognized repositories. The chosen repositories are:

a) IEEE-Xplore
b) ACM Digital Library
c) Science Direct (Elsevier only)
d) Wiley Online Library
e) Springer

The justification behind the selection of these online databases is that IEEE Xplore offers a number of important conference articles and symposium articles, while ACM Digital Library provides more articles from workshops used for the primary studies. The remaining repositories are equally important as they host journal articles that are related to test case prioritization. There are also important journals extracted from IEEE Xplore and ACM Digital Library.

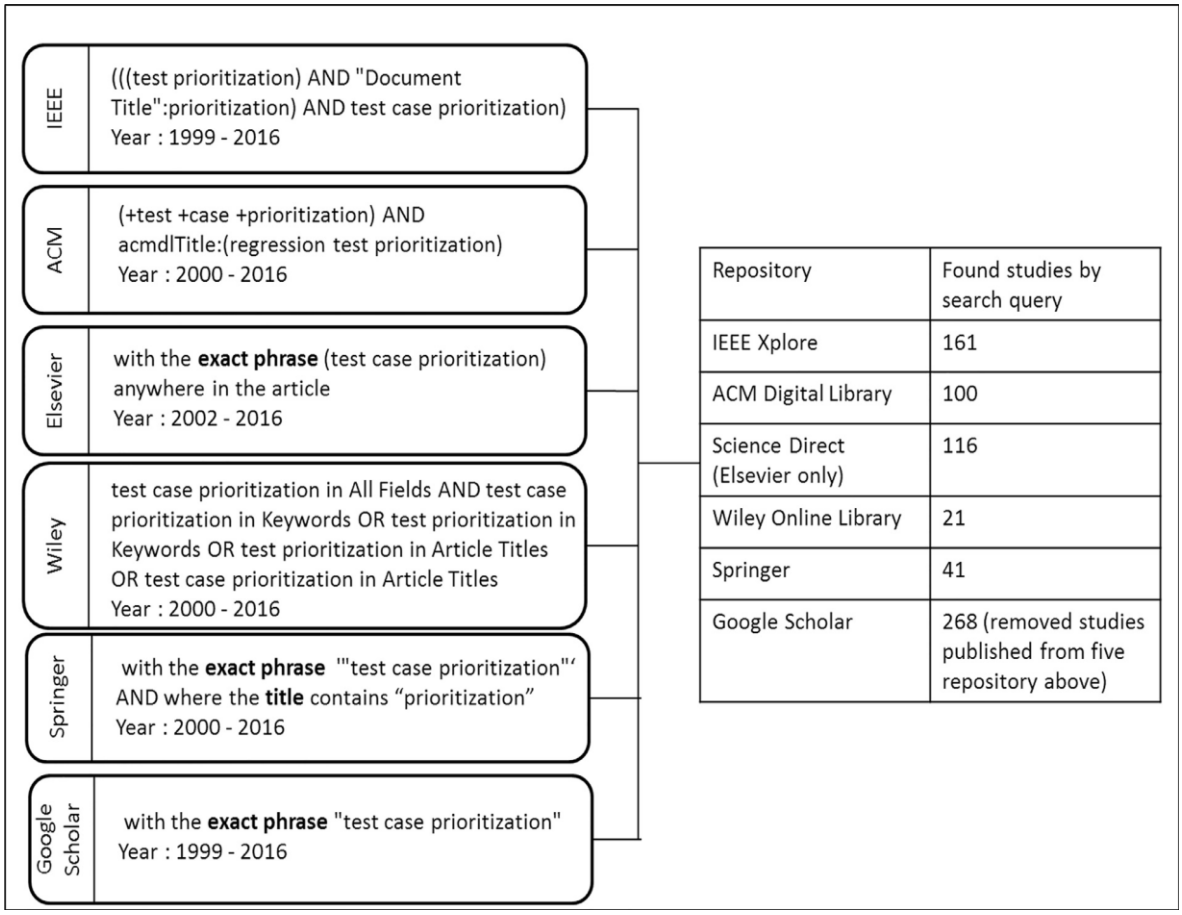
D. Search string identification

SLS is a well-known review technique for reviewing the literature with an extensive search aspect of the subject in the discussion from all relevant sources. Therefore, a systemic method to formulate search keywords in this SLS consists of the following steps:

a) Determination of significant terms based on RQs.
b) Determination of equivalent words for significant terms.
c) Determination of keywords in applicable studies.
d) Usage of the Boolean 'OR' and 'AND' operators as an alternative link between terms.

As the focus is to examine related studies in regression testing, TCP area to be precise, the results from previous studies are utilized in order to determine significant studies. The author's intentionally used test case prioritization as the exact phrase in most of the search queries since there are numerous research works that are related to testing test cases in regression testing.

As illustrated in Fig.2, different search strings were used in each repository. With a specific goal to retrieve conceivable significant reviews, the authors utilized these terms: "test prioritization", "regression testing", "test case prioritization", and "regression test prioritization". It ought to be noticed that if the authors utilized an exact phrase such as "test prioritization" alone, it will return an excessive number of unimportant reviews. Due to this, 'AND' operator was used to link "test prioritization" phrase and "test case prioritization" phrase to incorporate them into an alternative search term. As a result, quite a number of relevant studies were extracted from the combination of the phrases. To refine search outcomes, the authors used 'OR' operator for the phrases that appear in document titles and author keywords. The range of years published was set starting from 1999 until 2016.



Repository	Found studies by search query
IEEE Xplore	161
ACM Digital Library	100
Science Direct (Elsevier only)	116
Wiley Online Library	21
Springer	41
Google Scholar	268 (removed studies published from five repository above)

Fig2. Search strings for selecting studies from all repositories.

E. Study selection process

As mentioned in the previous section, SLS requires being conducted in an appropriate manner in order to produce a high impact review to the research domain itself. The SLS search process was initiated during the selection of the repositories. The first process was to identify several popular repositories in TCP research area. The next stage was a search stage, where, an exhaustive search was performed on all six selected repositories and all the prospective papers were assembled together to ease the selection process. Then, all the prospective papers under-went a selection stage in order to choose for relevant papers that were going to be used in the primary studies. Fig. 3 illustrates the search and selection process conducted for this SLS.

From Fig. 3, potential studies were identified from the search stage. To narrow down on the number of the papers to be reviewed, all the prospective studies were required to be justified to get the most significant studies. First of all, the prospective studies were required to go through inclusion and exclusion criteria. This process was essential to remove duplicates and unrelated studies. A detailed overview of the inclusion and exclusion criteria utilized for scrutiny is shown in Table 3. The first criteria to be satisfied were papers printed and issued in English only will be chosen. Studies that were not available in English were removed. Then, their abstract was briefly studied. The paper that does not have any association with research questions were excluded from the major studies list. For duplicated papers that appeared in different copies, the most recent ones would be the most completed and improved copies. They were selected while the others were removed.

After the inclusion and exclusion stage, quality assessment was applied. The quality evaluation of the chosen studies was accomplished by utilizing a weighting approach to examine significant studies that are adequate enough to answer all the RQs. The authors articulated five assessment questions shown in Table 4 to evaluate the comprehensiveness, reliability, and applicability of the nominated papers. Three optional answers with their respective score were given for each question: "Yes"=1, "Partly"=0.5 and "No"=0. Subsequently, various papers were rejected from this quality assessment. Consequently, only 80 papers were chosen for the primary studies. Upon the completion of this selection stage, 80 studies were identified to manifest the capability to answer all of the research questions derived earlier.

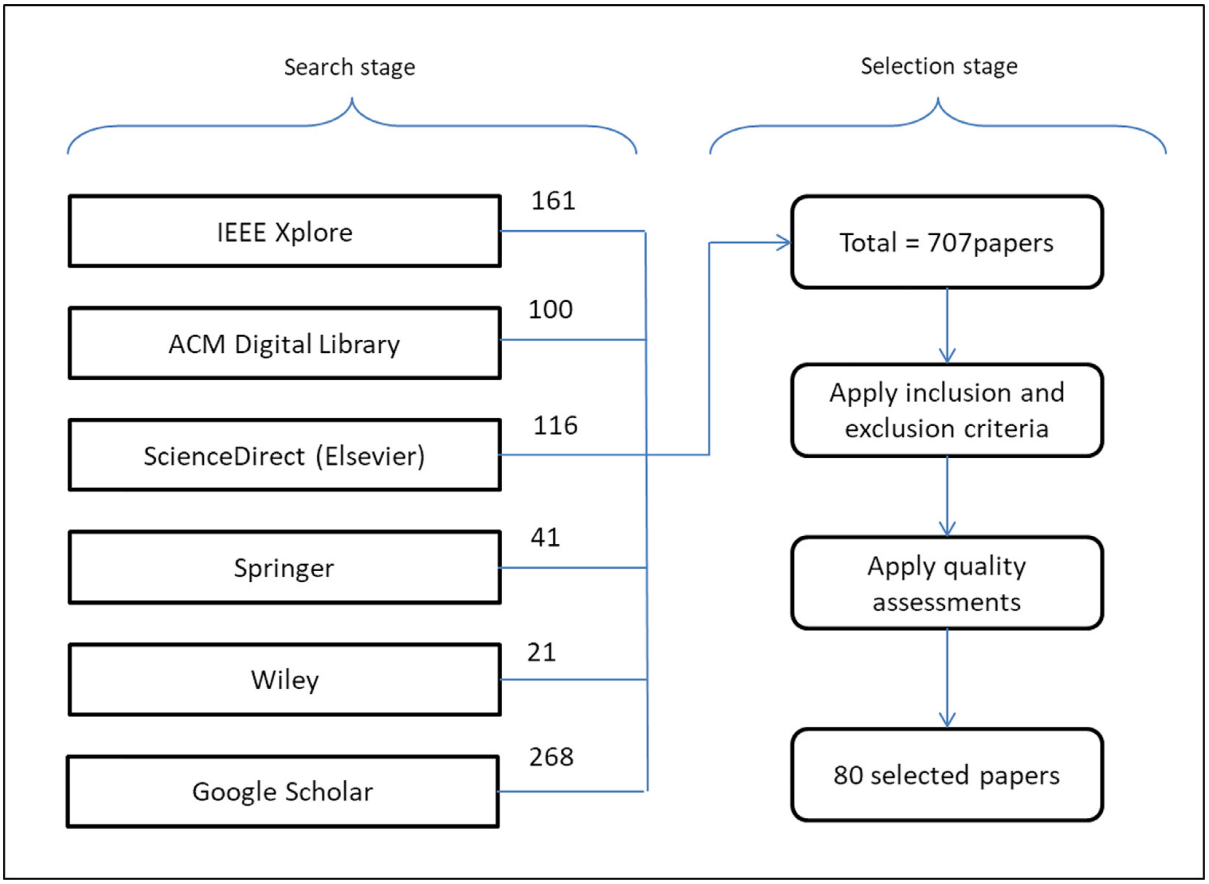


Fig3. Search and Selection Process

Inclusion criteria	Exclusion criteria
English as publication language	Non-English as publication language
Paper focusing on test case prioritization approaches	Paper does not have any relation with TCP approaches
Paper with complete bibliographic information from 1999 to 2001	Paper without bibliographic information
Paper is able to correspond to at least one research question	Identical studies (latest paper is included)

Table 3 Inclusion and exclusion criteria.

No	Question
1	Were the research objectives stated precisely?
2	Were the planned approaches stated clearly?
3	Was the experimental strategy appropriately designed?
4	Did the experiment apply on a case study or a controlled experiment?
5	Does the exploration enhance the scholarly world?

Table4. Quality assessment questions

F. Data synthesis and extraction method

The principle of data synthesis is to simplify evidence presentation from the nominated papers to ease data extraction process. This is in order to answer all of the research questions. 80 selected primary studies underwent an additional inspection with respect to the content assessment measures as shown in Table 5 to determine the Specified matters for each paper.

The point of this measure is to synchronize the primary studies and enhance the assessed papers for clarity. This will help for the purpose of data extraction from papers which response exactly to the research questions. Brief important types of data showing the mapping of synthesized data to research questions are shown in Table 6 below.

Nominated paper	Description
Paper references	Paper title, publication year, and sources
Type of paper	Journal article, conference proceeding, symposiums, and workshop
Paper focus	Main ideas, complications, inspiration and purpose
Research methodology	Case study, experimentation, reviews, and literature surveys
Application domain	Depiction of the specific situation
Constraints	Study's limitations for future improvement

Table5. Contents assessment measures

Research questions (RQs)		Type of data extracted
RQ 1	RQ 1.1	Types of regression testing techniques, studies, and bibliographic references.
	RQ 1.2	Types of TCP approaches solution types used to conduct the techniques.
RQ 2	RQ 2.1	Brief description and test purpose of each approach, formal hypothesis, test strategy, threats to validity.
	RQ 2.2	Formal hypothesis, advantages, and a threat to validity.
RQ 3	RQ 1	Tool/environment for experimental setup, data collection method, process, and evidence type or measure for the effectiveness of TCP approaches.
RQ 4	RQ 1	Evaluation metrics and soft-ware artefact type.

Table6. Data collection for research questions framed.

IV. RESULT AND DISCUSSION

This section outlines the results with respect to the research questions. The summary of the primary studies was presented first, followed by each research question, answered in different sub- section.

A. What is the research trend of TCP techniques in regression testing? (RQ 1)

As prioritization on test cases had only gone through test case selection in early studies [3] , TCP was then suggested and assessed in a broader context. The first aspect of RQ1 was to determine the current trend of TCP techniques' studies. Referring to Fig5, the number of papers published through the year shows a positive increment beginning from 1999 until 2016. It can be concluded that TCP has been recognized as an important element in regression testing among researchers. In real world scenarios, it is quite hard to realize which tests will actually detect faults. Hence, that is why test case prioritization approach needs to have other approach backups, expecting that a certain number of backup approaches will end in boosting fault discovery in different ways. There are many test case prioritization approaches which have been proposed by researchers. As many as eight broad approaches were described by Singh. All these TCP techniques approaches were originally grouped based on some commonalities such as phases selected, available resources (requirement, test cases), and desired output (time execution, APFD). For example, Requirement-based TCP can be initiated during or at the end of requirement phase, by using the requirement resource itself. Fig 6, shows other discovered approaches.

As shown in Fig 6, the authors grouped the TCP approaches into seven main dimensions, which seem to be popular among researchers. These categories are reported in recent mapping studies and they have various

common characteristics [14]. However, the authors combined some approaches under 'others-based' dimension, which were not really popular among researchers including topic model based [18], multi-criteria based [19, 20]. Each approach presents different process and dataset in performing regression testing.

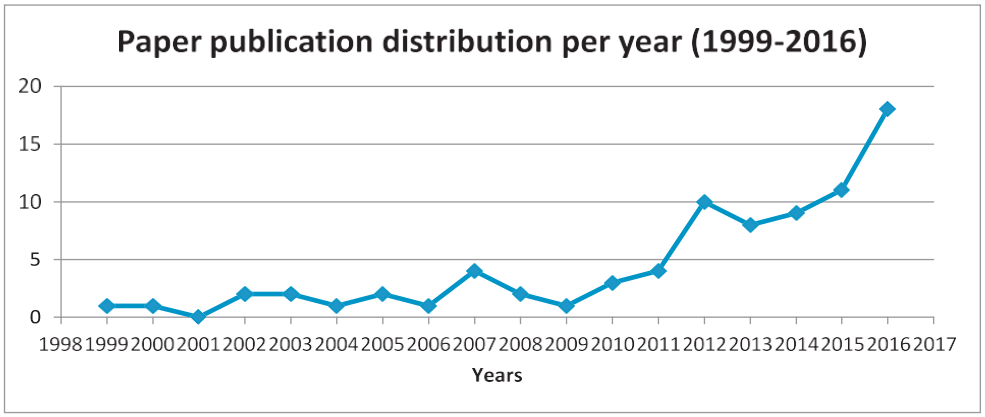


Fig5. Number of papers published per year

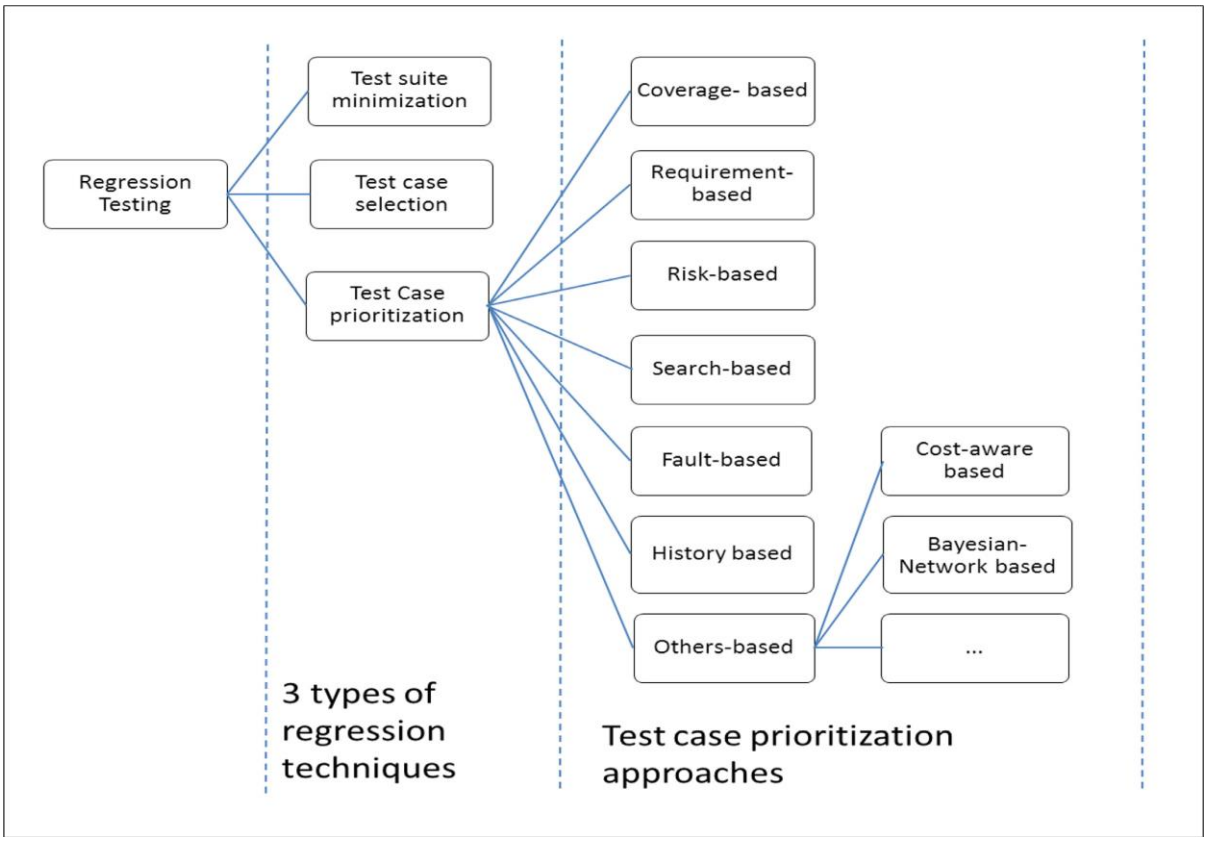


Fig6. The taxonomy of TCP

B. What are the processes involved in TCP? (RQ 3)

Soft-ware engineering highly concerns on how the engineering processes are applied to soft-ware development in a systematic way. Therefore, it is necessary to have this RQ to be investigated. In order to answer this question, the primary studies were examined further into their experiment flow. Every experiment should have their own process, however, in this SLS; processes involved in TCP are only highlighted from several studies. The reason be- hind this is because, numerous works exhibit similar process flow, with the only notable difference in terms of the addition of one extra step to an existing process flow. From the primary studies, the authors determine 19 processes proposed by several researchers. All these 19 works were selected as their processes were clearly presented. There were also other clearly stated processes in other studies, but they

were not covered by the RQ's. Based on the 19 processes realized, the authors illustrate the basic flow of TCP process as shown in Fig 9.

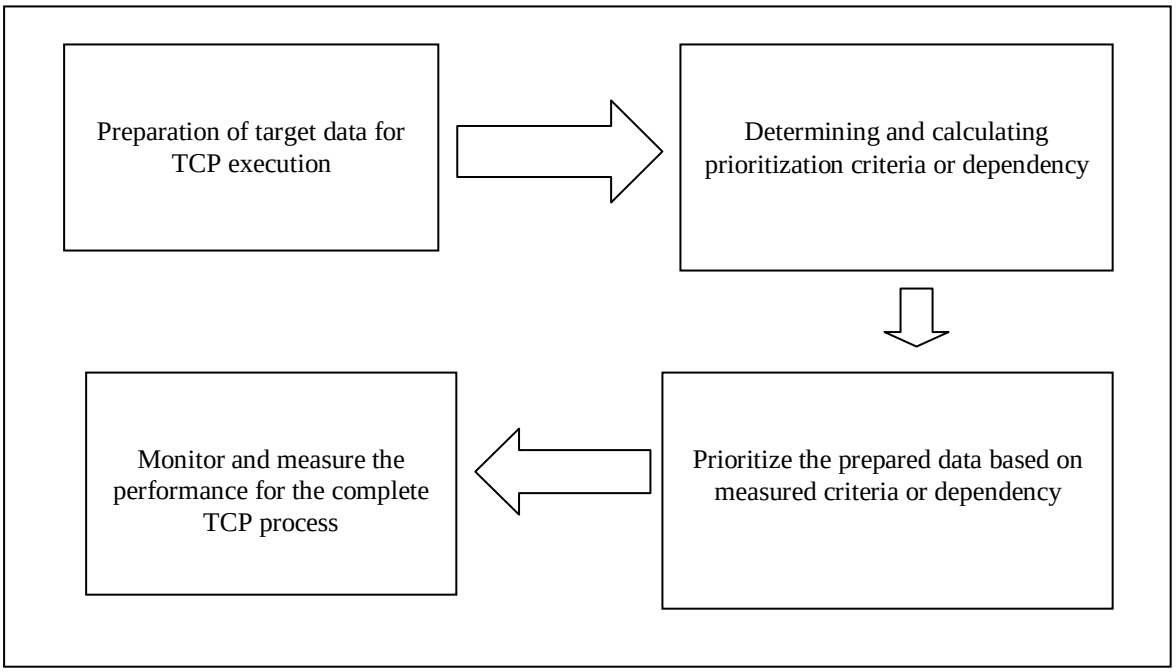


Fig9. Standard flow of TCP technique

As shown in Fig 9, TCP process starts with the preparation of targeted data. Even though there is no single paper stating clearly that the TCP process starts with the preparation of data, it is compulsory for any experiment or research to identify which information or data that will be used. The data or information in TCP can be in the form of requirement statement, system models, and source code. The process is followed by, determining and calculating prioritization criteria or dependency based on the data chosen, first, the test cases that have most coverage and which are unique in coverage-based approach. While in risk-based approach, the risk value related to each test case was defined and calculated before any prioritization was performed. After prioritization criteria determination stage, next is prioritization process. Test case prioritization activity can only be started after the criteria have been determined or measured, prioritization of test cases start after the requirement criteria such as customer priority is determined and calculated, TCP process can be repetitive until all requirements, test cases, or prescribed faults are fully covered or satisfied. Finally, the result is monitored and the performance of a complete TCP process is measured. However, each approach may start this process in a different timeline within a project schedule based on its chosen data. For requirement statement data, TCP can start during or right after requirement phase. For system models' data, it starts right after the design phase is completed and source code needs to wait until it is available.

C. What is the evaluation metric used in TCP along with the soft-ware artefact? (RQ 4)

To answer this research question, this section is divided into two sub-sections, comprising evaluation metric and soft-ware artefact sub-sections, with the relationship among them.

C.1. Evaluation metrics

It is essential for any approaches proposed in test case prioritization to perform metric measurement to assess their effective-ness. Evaluation metric is important to measure the efficacy of any proposed TCP approaches in prioritizing test cases and to bench- mark its effectiveness against other existing approaches.

The most widely used metric is APFD with a 51% distribution, followed by Coverage Effectiveness (CE) 10%, APFDc (APFD with cost consideration) 9%, time execution 7%, and others 23%. Average Percentage of Faults Detected (APFD) is evidently the most utilized metric favoured by researchers in the primary studies. This metric was originally introduced by [3] in early TCP research, and later used massively by other researchers. APFD is a metric used to quantify on how quick an arranged and optimized test suite can discover defects. The result of APFD ranges from zero to 100, where a greater value indicates a better fault revealing rate. From this metric, researchers have come up with other metric that evolves APFD metric by considering other factors such as cost or time constraints to fulfil different objectives of prioritization.

Coverage effectiveness (CE) comes as the second most utilized metric. CE is a metric that integrates the size of test suites and the coverage of each test case. It is the ratio between the size of the whole test suite

and the coverage of reordered test suite that reveals all faults or meets all requirements. CE values range from zero to one, where a higher value indicates a better effectiveness in coverage. Next is APFDc with a 9% distribution, which is an APFD comparable metric with an inclusion of a cost factor. APFDc has been utilized by researchers, to measure the effectiveness of fault detection rate over cost. This metric is the most widely used metric in history and fault-based approaches, since history-based concerns on improvising previous testing that includes cost factor, while fault-based focuses on prescribing fault that are most likely to be the costly one.

Time execution metric is used by 7% of the distribution of re- searchers in this SLS. Time execution metric is primarily used in search-based TCP approach to verify the effectiveness of a proposed algorithm in reducing TCP time. Finally, other metric comprises a 23% distribution, which includes several types of metric available. Most of the metrics are adapted from APFD, which has been turned into ASFD, WPFD, TSFD, APBC, APDC, APSC, and NAPFD to answer different TCP objectives.

To further support the significance of all evaluation metrics, ANOVA test has been utilized by several researchers. Analysis of variance (ANOVA), is a statistical method. ANOVA is a statistical test that aims to evaluate the difference of the means of three or more groups. The groups' data must be in numerical type. ANOVA hypothesizes the findings with both null hypothesis and alternative hypothesis. A null hypothesis is where all treatments are having equal values while alternative rejects the null hypothesis. ANOVA test in TCP has been used to verify the impact of the proposed work by measuring a statistical significance of the investigated approach, by comparing any metrics, such as APFD or execution time.

C.2. Artefact

An artefact is required in completing a controlled research in testing practices. Artefacts can be in the form of test cases, soft- ware, coverage information, soft-ware requirements, fault records, history evidence, and others. The utilization of artefact depends on the type of research and its suitability to the experiment. A careful examination of the artefacts utilized by different TCP approaches is presented by Singh [13]. The study highlighted a number of artefacts assessed by researchers. In this SLR, the distribution of artefacts utilized is presented along with their usage in different TCP approaches. Fig. 11 shows the TCP-related artefacts assessed by re- searchers.

On the subject of the artefacts assessed, our results conform here a 50% distribution of the artefacts are accessible with no restriction from Soft-ware Infrastructure. In this SLS, a 36% distribution of the soft-ware artefacts has been extracted from SIS, as illustrated in Fig. 11. Meanwhile, real case study has been used in several studies, which is considered as an artefact in this SLS. The artefacts from SIS are mainly divided into three types comprising, siemen, space, and TLS. Siemen set consists of seven benchmark program which is written C programming language.

The next popular SIS program among researchers is Test Specification Language (TSL) with an 11% distribution. This category is named TSL after a TSL tool that is used in executing the program. TSL program consists of four programs which are flex, grep, gzip, and make, where all of them are UNIX utilities. The last SIS is a 'Space' program with 9564 lines of code (6218 executables), which makes up 8% of the total artefacts assessed. The space program has been used to validate the content within a file and functions as an interpreter for an array definition language (ADL) grammar. In short, SIS provides information for each of C object, excluding requirement specification documentation. Over- all, SIS programs have been widely utilized in TCP approach, for studies that are not related to requirements.

We analysed that 68% of utilization of SIS programs have been utilized under search-based TCP approach, while 16% of the studies have utilized SIS programs in fault-based TCP . Other approaches that utilize SIS include, coverage- based 4%, history 2%, and others 10%. The reason behind this distribution of studies concerning utilization SIS is because SIS only provides artefacts in the form of source code, test suites, fault in- formation, coverage information, and some history evidence. An artefact such as requirement is not provided, which explains why requirement-based TCP does not utilize SIS programs.

V. RESEARCH FINDING

In soft-ware development, prioritizing test case is an essential activity in testing phase [3} . At the point when client hopes are high, promised time is short, assets are constrained, and the developed soft-ware must have the capability to fulfil the soft-ware requirements with fewer faults, TCP technique is beneficial to reduce the cost, time, and soft-ware fault since TCP itself concerns to order test cases for early optimization in fault detection. A de-tailed approach of TCP is very important, not only to be used to optimize the test case execution, but also to aid project manager to organize project deliveries, and maybe to make some necessary adjustment. Therefore, the TCP impact in regression testing must be emphasized more.

In answering RQ1, all of the primary studies were used to answer this research question. As for the result, it can be said that TCP approaches are still broadly open for improvements. There were positive increments of TCP publication starting from 1999 until 2016. New approaches in TCP are introduced constantly

every year, and implementation of artificial intelligence element has been a trend among researchers in the latest study. However, other approaches have their own supporters, with their own advantages. For example, in other approaches, a multi-objective or criteria technique has shown quite a number of supporters as it has the capability to tackle two or more different kinds of objectives in one prioritization. This approach can also be easily combined with other techniques, which makes it a more interesting and promising approach. Therefore, it can be concluded that TCP is recognized as an important element in regression testing among researchers currently, as it has the capability to increase the effectiveness of testing in terms of fault detection rate, cost, and time.

For RQ2, 42% of the primary studies were examined thoroughly. As a conclusion, each approach has specified potential values, ad- vantages, and limitation. The inputs and dataset type play an important role in the determination of their advantages and limitation. For example, the requirement-based approach uses customer priority during requirement elicitation as inputs to prioritize and generate a test case. Risk-based may also use requirement risk as one of the inputs to execute prioritization process. This indicates that both may have their own advantages against other approaches in terms of TCP execution starting point since both may start prior to code availability. The differences in term of strengths and limitations of these approaches are required to give a hint on how TCP approaches function and serve as a motivation for any changes in the future.

For RQ3, 19 out of 80 primary studies were evaluated regarding their experimental setup. The implementation of TCP process shows a significant role in certain project (RQ3). TCP techniques benefit project managers in adjusting their project schedules in order to counter the constraint that exists within the project development process. Variations in the starting point of TCP pro-cess among the approaches provide a different timeline and benefit to project manager to choose which approaches suite with the project schedule and available resources.

For RQ4, all of 80 primary studies have been examined thoroughly. The evaluation metrics utilized in these primary studies with the reasons behind their creations have been covered. APFD metric remains as the main metric used in all TCP approaches. However, there is quite a number of new metrics that have been introduced, which is adapted from APFD to support different objectives in different studies. As for the artefacts, the evidence sug-gests that programs extracted from SIR remains the most popular choice among researchers. However, SIR does not support requirement and model based TCP approaches.

VI. THREATS TO VALIDITY

This SLS has some limitations recognized that may threaten its validity. Similar to previous reviews, the potential threats of this systematic review are associated with an imperfect collection of primary studies and imprecise data synthesis and derivation. However, the authors cannot completely ensure that the authors have gained all the accessible reviews related to TCP to answer all RQs. There are two possible issues. Firstly, a possible issue is the difficulty in finding appropriate search strings for different repositories. However, there might be other significant studies which may use other keywords. Secondly, some significant or related studies might have existed in non-English publication. It misses out the relevant non-English studies although it cannot be avoided.

VI.1. Imprecise data synthesis and extraction

Imprecise data synthesis and fragmentary extraction process from the primary studies could be the next potential threat to the validity of this SLS. This may be due to unsystematic data extraction or invalid classification of data. To reduce this problem, manual scrutiny was applied to reduce the possibility of inaccurate data extraction by focusing on the data elements collected and, a set of specific quality assessments were applied to avoid inaccurate inclusion of desired studies.

VI.2. TCP related field

Within regression testing, there are three techniques comprising, TCP, TCS, and TSM. Test case selection (TCS) is almost similar to TCP as the idea behind TCS is to localize fault, select related test cases, and execute it as a suite. However, in this SLR, the authors did not discuss the current trend of TCP in supporting fault localization issues which may be a potential threat to the validity of this SLR. TCP does support in fault localization. A TCP approach may also be turned into TCS after prioritization is completed and selection can be made based on prioritization result. Work by proposed a technique known as diversity maximization speedup (DMS), which ranks up program unit while performing a fault localization for TCP and the results demonstrated that the technique was able to realize the aimed cost.

VII. CONCLUSION

As the purpose of this study was to classify and criticize the current state and trend of test case prioritization approaches, SLS scheme was used to conduct this study. With this SLR scheme, some applicable RQs were formulated according to the aim of this study. This research was conducted through finding, classifying, evaluating, and understanding all of the primary studies. The motivation of this review was to discover any areas or fields that are likely to undergo any sort of improvement via systematic assessment of applicable and important studies in TCP approaches.

During the review process, the appropriate primary studies were recognized and evaluated. The data pulled out from the studies were synchronized. The authors structured the outcomes into tables and figures, which are intended to ease the understanding among the different research groups that work on the same TCP area. From the study, it was discovered that:

1. There were quite a number of prioritization techniques that exist and improvements are still required.
2. AI application is quite popular since it can be used for different problems and the empirical data availability for AI experimental setup is high.
3. The data or information classification in TCP such as requirement statement, system models, and source code needs to be investigated further in future.
4. The specific time frame for the execution process of TCP for each approach needs to be detailed out.
5. Early prioritization in the early stage of a system development life cycle is worth to be investigated further to ease project manager and development team in making the necessary adjustment.
6. Human involvement in decision-making and estimation needs to be changed into a computerized and automated reasoning to reduce human error.
7. How does TCP support fault localization compared to TCS? This issue could be an interesting point to be discussed in the next work.

REFERENCES

[1] P. Ralph, Soft-ware engineering process theory: a multi-method comparison of sensemaking-coevolution-implementation theory and function-behav structure theory, Inf. Softw. Technol. 70 (2016) 232–250.
[2] G.J. Myers , T.M. Thomas , C. Sandler , The Art of Soft-ware Testing, vol. 1, John Wiley & Sons, 2004 .
[3] G. Rothermel , R.H. Untch , C.C. Chu , M.J. Harrold , Test case prioritization: an empirical study, in: Soft-ware Maintenance, 1999. (ICSM '99) Proceedings. IEEE International Conference on, 1999, pp. 179–188.
[4] H.K.N. Leung, Insights into regression testing, in: Proceedings of the International Conference on Soft-ware Maintenance, 1989, pp. 60–69.
[5] S. Elbaum , G. Rothermel , J. Penix , Techniques for improving regression testing in continuous integration development environments, in: Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Soft-ware Engineering –FSE 2014, 2014, pp. 235–245 .
[6] P.K. Chittimalli , M.J. Harrold , Recomputing coverage information to assist regression testing, IEEE Trans. Softw. Eng. 35 (4) (2009) 452–469.
[7] S. Yoo , M. Harman , Regression testing minimisation, selection and prioritisation : a survey, *Test Verification Reliab.* 0 (2007) 1–7.
[8] D. Jeffrey , N. Gupta , Improving fault detection capability by selectively retaining test cases during test suite reduction, IEEE Trans. Softw. Eng. 33 (2) (Feb. 2007) 108–123.
[9] S. Elbaum , P. Kallakuri , A.G. Malishevsky , G. Rothermel , S. Kanduri , Understanding the effects of changes on the cost-effectiveness of regression testing techniques, J. Softw. Test., Verification Reliab. 12 (2) (2003) 65–83.
[10] S. Elbaum , A.G. Malishevsky , G. Rothermel , Test case prioritization: a family of empirical studies, IEEE Trans. Softw. Eng. 28 (2) (2002) 159–182.
[11] B. Kitchenham , in: Procedures for Performing Systematic Reviews, 33, no. TR/SE-0401, Keele University, Keele, UK, 2004, p. 28.
[12] B. Kitchenham , O.P. Brereton , D. B. , et al. , Systematic literature reviews in soft- ware engineering –a systematic literature review, Inf. Softw. Technol. 51 (1) (2009) 7–15.
[13] Y. Singh, Systematic literature review on regression test prioritization techniques difference between literature review and systematic literature, Informatics 36 (2012) 379–408.
[14] C. Catal, D. Mishra , Test case prioritization: a systematic mapping study, Softw. Qual. J. 21 (3) (2012) 445–478.
[15] A. Kumar, K. Singh , A Literature Survey on Test Case Prioritization, Compusoft, 2014 .
[16] P. Kiran , K. Chandraprakash , A literature survey on TCP-test case prioritization using the RT-regression techniques, Global J. (2015).

[17] P. Achimugu , A. Selamat , R. Ibrahim , M. Naz , A systematic literature review of soft-ware requirements prioritization research, Inf. Softw. 56 (2014) 568–585.
[18] S. Thomas , H. Hemmati , A. Hassan, Static test case prioritization using topic models, Soft-ware Engineering, 2014.
[19] S. Sampath , R. Bryce , A.M. Memon , A uniform representation of hybrid criteria for regression testing, IEEE Trans. Softw. Eng. 39 (10) (2013) 1326–1344.
[20] A.B. Sanchez, S. Segura, A. Ruiz-Cortes, A comparison of test case prioritization criteria for soft-ware product lines, in: Soft-ware Testing, Verification and Validation (ICST), 2014 IEEE Seventh International Conference, 2014, pp. 41–50.

