Test case prioritization techniques in software regression testing: An overview

Muhammad Qasim 1, Asifa Bibi 2, Syed Jawad Hussain 3, N. Z. Jhanji 4, Mamoona Humayun 5, Najm Us Sama 6, *

1Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Karachi, Pakistan
2Department of Computer Science, University of Northern Virginia, Annandale, USA
3Department of Computer Science, Barani Institute of Information Technology, Pir Mehr Ali Shah Arid Agriculture University, Rawalpindi, Pakistan
4School of Computer Science and Engineering, Taylor’s University, Subang Jaya, Malaysia
5Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakakah, Saudi Arabia
6Department of Computer Science, Deanship of Common First Year, Jouf University, Sakaka, Saudi Arabia

1. Introduction

Software Testing (ST) plays an important role in the Software Development Life Cycle (SDLC) as it guarantees defect-free software and customer satisfaction. Despite the complexity and importance of this phase when it comes to testing a small amount of time has been left behind which makes this phase costlier. ST techniques involve the execution of software to find if all the components are as per customer need. If any issue is found, it will be considered a defect or bug.

There are lots of stages of ST and lots of ST techniques are used. Among all the ST techniques the most important one is Regression Testing (RT). The importance of RT could never be ignored as this verifies that newly updated code impacting previously build functionality or not? It doesn’t matter how many times we make changes in the code we have to run RT to make sure software stability against the code change.

Because of the limited time given for RT, it becomes the difficult phase in the sense of how and which test case to be selected for RT suits. Different methodologies have been used to make the best test suite for regression testing which is called Test Case Prioritization (TCP) or Test Case Selection (TCS). These methodologies include search, coverage, fault, requirement, history, risk and cost-effective based, etc. (Shah et al., 2019). Each of the methodologies has its own algorithms. Best TCP/TCS is dependent upon different criteria like the selection of dataset, incomplete data extraction (Shah et al., 2016). In this study, we have selected the most recent algorithms...
used for software TCP from the most popular repositories like IEEE Access, Elsevier, and Springer.

In this survey, we have studied 40 articles from 2010 to 2020 on TCP. This study organized as follows. In Section 2 which is a literature review, we will describe all the chosen algorithms in Table 1. In Section 3 we will compare the results generated by using each algorithm. In Section 4 we will describe research methodologies and try to answer the research questions. Section 5 contains state-of-the-art challenges and Section 6 contains conclusions and future research topics.

<table>
<thead>
<tr>
<th>S.No.</th>
<th>Name of Algorithm</th>
<th>Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Firefly Algorithm</td>
<td>Firefly Algorithm</td>
</tr>
<tr>
<td>2</td>
<td>Genetic Algorithm (GA)</td>
<td>Optimal FA</td>
</tr>
<tr>
<td>3</td>
<td>Ant Colony Optimization</td>
<td>Hypervolume GA</td>
</tr>
<tr>
<td></td>
<td>Algorithm (ACO)</td>
<td>ACO</td>
</tr>
<tr>
<td>4</td>
<td>Local Beam Search (LBS)</td>
<td>Modified ACO</td>
</tr>
<tr>
<td>5</td>
<td>Particle Swarm Optimization (PSO)</td>
<td>Weight Hybrid PSO</td>
</tr>
<tr>
<td>6</td>
<td>Greedy Algorithm (GA)</td>
<td>Greedy PSO</td>
</tr>
</tbody>
</table>

2. Literature review

Test Case Prioritization (TCP) not only saves time and resources but also stakeholder confidence if we used it properly. Various methods and techniques have been used in this regard to obtain the best optimal test suite for Regression Testing. In this study, we will cover and summarize all those algorithms as shown in Fig. 1.

3. Firefly algorithm

Firefly Algorithm (FA) after getting inspired by the flashy behavior of the fireflies (Shah et al., 2019).

- Fireflies are attracted to each other because all of them are unisexual.
- The attraction of fireflies is depending upon their brightness. Both are directly proportional to each other.
- Fireflies only move randomly if the same/no brightness is found nearby.

Quite a few algorithms were proposed using the Firefly technique.

3.1. Optimal firefly algorithm

This algorithm uses a metaheuristic firefly algorithm to generate an optimal path. In this modified firefly algorithm, he used a new objective function/brightness function and guidance matrix to traverse the graph (Shah et al., 2016).

The process starts with creating Control flow graphs (CFG) and Data flow diagrams (DFD) and generating XML files from the CFG diagram. This XML file goes as input into Optimal Firefly Test Sequence Generator (OFTSG). OFTSG perform the following steps (Shah et al., 2016):

- Calculate nodes and edges
- Calculate Cyclomatic Complexity
- Create Adjacency matrix which is the edge between the nodes is considered as 1 everything else is 0.
- Create Guidance matrix using Eq. 1
- Path traversal: Algorithm generates best path sequences.
- Path Prioritization
- Calculation of brightness value at every node using the Eq. 2
- Calculation of mean value of brightness.

Higher the mean value of brightness will have higher priority.

\[ GF = 10 \times \left[ \frac{CC_i \times \left( (N - i) - 0.1 \right)}{\left[ \frac{1}{CC_i + \text{rand}(j)} \right]} \right] \]  \hspace{1cm} (1)

\[ f(x) = \left[ \frac{1000}{\left[ \frac{1}{CC_i + \text{rand}(j)} \right]} \right] \]  \hspace{1cm} (2)

The author tested this algorithm using 4 different programs as shown in Table 2.
This shows that as the number of states increases the cyclomatic complexity increases which causes an
increase in the percentage of redundancy in path coverage. Results of this study were compared with the state of the art algorithm Ant Colony Optimization Algorithm (ACO). This shows FA generates better paths with reducing or no redundancy as compared to ACO as shown in Table 3.

Following limitations were found during this study, first, the proposed algorithm was not tested on any benchmarks program with a big test case pool. Secondly, it’s not clear that which testing criteria are used for example statement coverage, fault coverage, etc.

Panthi and Mohapatra (2015) used the same methodology to minimized test cases by exploring the state diagrams. Nor Hashim and Dawood (2018) proposed a similar approach to minimized test cases and the process has been analyzed using a UML statechart diagram. This study also shows prominent results in the form of minimizing the number of test cases.

Both used ATM case studies to test their approaches which provide satisfactory results.

3.2. Firefly algorithm

Khatibsyarbini et al. (2019) used Firefly Algorithm with string metrics as a fitness function to find the best TCP arrangement.

Steps of Firefly Algorithm:

- Extraction of test cases from dataset
- Calculating distance and weight of test cases
- Perform movement update based on highest values of weight/distance
- Selection of test cases based on shortest distance as an optimal sequence.

Two variables Term Distance Inverse Document Frequency (TFIDF) and Term Frequency Inverse Document Frequency (TFIDF) were used to store distance among the test case and frequency (weight) of the test case respectively. The weight of test cases could be calculated as follows given in Eq. 3 (Khatibsyarbini et al., 2019).

\[
TFIDF = \frac{t \times \log N}{nt}
\]

The next move decided by the highest value of weight over distance and the shortest distance among test cases was selected as the optimum test case sequence. FA uses average percentage fault detection rate (APFD) to calculate the rate of fault detection which could be calculated using the formula given in Eq. 4 (Khatibsyarbini et al., 2019).

\[
APFD = 1 - \frac{RA + TF + \ldots + TF + \frac{1}{2}}{nm}
\]

Six Software-artifact Infrastructure Repository (SIR) programs were used for testing. Three of them are UNIX-based programs (flex, grep, gzip) and three are SIEMENS programs (tcas, cs-tcas, j-cas) as shown in Table 2. Comparisons were made in this approach with state-of-the-art algorithms PSO, LBS, Greedy, and GA and among all of them, FA shows prominent results although LBS has quite better results at some places as shown in Table 3. Following limitations were observed during this study. The first one is a selection of datasets. The dataset used for testing this approach is very small SIR benchmark programs. This approach may give different results when applied to a bigger size of test case spools.

The second is less data utilization. For example, only distance and weight of test cases were used as input while we have a lot of other data available like system coding, etc. that could give totally different results. The third and final one is that the only TCP techniques were considered on the other hand we have a lot of other similar techniques available which was not included in validation.

3.3. Hybrid firefly algorithm (HFA)

Su et al. (2020) proposed this hybrid model of the firefly algorithm, main purpose is to prioritize test cases and reduce the test cost. HFA adopted a hybrid model in which first it correlates all the available test data and test cases and secondly applies enhanced FA on this hybrid model. The new fitness function of this HFA is shown in Eq. 5.

\[
f(x_{ij}) = \max_{k} \left\{ \frac{W_{i_{k}}}{\text{Random}_{i_{k}}}, \text{Brightness}_{i_{j}} \right\} \quad j - 1 \in \text{Order}, i \text{ unOrder}
\]

where, Wi-1 shows the test cases weight i-1, a label “Order” is used for the test cases where the priority is known, “un-order” is for test cases where the priority is unknown, Random (·)=\{-1,0,1\}, N is the total number of test cases.

The steps of this algorithm are as follows (Su et al., 2020):

- Define the new fitness function as shown in Eq. 5.
- Calculate brightness of firefly using Eq. 6.
- Calculate the degree of attraction by creating a distance matrix as shown in Eq. 7.
- The movements of the fireflies are based on brightness only. Once all nodes have been visited, the firefly stop moving and their path is recorded.
- In the end, the smallest path will be an optimal sequence.

\[
\text{Brightness}_{i_{j}} = W_j \left( \frac{n_j}{n_j \times \log \left( \frac{n_j}{|d|} \right)} \right)
\]

\[
X_{i+1} = X_i + \beta \gamma X_i (X_i - X_j) + \alpha e^x
\]

where, \(\beta, \alpha, \gamma, \gamma, \) and \(e\) are constants, \(\beta\) is the light absorption rate, \(\gamma\) is a random factor of a uniform distribution, \(\gamma\) is attraction coefficient, HFA used three Software-artifact Infrastructure Repository (SIR) programs for testing. Which are flex, grep and gzip as shown in Table 2. Comparisons were made in this approach with state-of-the-art algorithms FA, PSO, and Greedy and among all of them, HFA gave the better execution time and best performance as shown in Table 3.
The following limitation was observed during this study the selection of the dataset. The dataset used in this approach is three very small SIR benchmark programs. This approach may give different results when applied to a bigger size of test case spools.

3.4. Hypervolume genetic algorithm (HGA)

The method uses GA with Hypervolume Indicator to prioritized test cases with more than three testing criteria. He proposed an enhanced genetic algorithm which is called HGA (Hypervolume-based Genetic Algorithm). Earlier GAs was used to solve only single-objective optimization problems. But hypervolume allows us to combine multiple-objects to test as a single object. HGA consist of two parts (Di Nucci et al., 2018; 2015).

Part-I
1. Initialization of cumulative coverage scores.
2. Computation of cumulative coverage scores (cost, branch, and statements) for each testing case.
3. Computation of actual coverage using equation 8.
4. Normalized Hypervolume (HP) using Eq. 9.

\[
I_B(t) = \sum_{i=1}^{n} [\text{cost}(p_{i+1}) - \text{cost}(p_i)] \times |\text{Cov}_i(p_i)| \times \ldots \times |\text{Cov}_m(p_0)|
\]

\[
I_{HP}(t) = \frac{\sum_{i=1}^{n} [\text{cost}(p_{i+1}) - \text{cost}(p_i)] \times |\text{Cov}_i(p_i)| \times \ldots \times |\text{Cov}_m(p_0)|}{\text{cost}(p_0) \times |\text{Cov}[max]| \times \ldots \times |\text{Cov}[min]|}
\]

Eq. 8, 9

Part-II
1. Test cases order generation.
2. Permutations evaluation by IHP (t) metric.
3. GA evolution derivation.

HGA used five Software-artifact Infrastructure Repository (SIR) programs for testing. Which are bash, flex, grep, gzip and sed as shown in Table 2.

Comparisons were made in HGA with state-of-the-art algorithms Additional Greedy Algorithm, GA (AUC metrics), NSGA-II, GDE3, MOEA/D-DE, and among all of them HGA shows the prominent results as shown in Table 3.

Following limitations were observed during the study of this algorithm. The first one is the type of construction used to validate the algorithm depends upon the results obtained. In order to validate the results APFDC matrix was used and only three testing criteria were used which include statement coverage, fault coverage, and execution cost. On the other hand, if we include the time parameter the results could be different. The second one is internal validity. a) Only different types of GA were used to compare the results of this algorithm. It would be highly appropriate to choose some other algorithm as well. b) Tuning of the parameter could really affect the validity of results Study shows that the value of parameters was kept the same as used in basic GA. Results should be tested by changing the values of parameters. The third which could affect the results is the dataset used in this approach which are six trademark programs from SIR only. The algorithm should be validated against real-time industry programs to test results.

3.5. Ant colony optimization (ACO)

ACO is another prominent approach used for test case prioritization. Following different types of the algorithm has been proposed based on ACO.

Ant colony optimization is a metaheuristic optimization methodology that was first introduced by Márquez et al. (2016). Zhang et al. (2019) used requirement-based criteria in their proposed ACO technique to solve the prioritization problem. He divided it into two ways TCP-ACA-1 in which he calculated the distance between two test cases and TCP-ACA-2 which is used to update the value of pheromone continuously (Zhang et al., 2019).

This method was successfully used in solving various optimization problems and the Test case prioritization problem is one of the combinatorial optimization problems where ACO showed impressive results. As per the technique used by Gao et al. (2015) during the test case prioritization problem, there are two rules that must be followed while using ACO methodology.

ACO algorithm works as follows: as the start ants are placed on the initial Test case list from where they randomly select the next node until all the faults are identified. It updates the pheromone value once each iteration is finished. The two most important steps in this algorithm are given below (Gao et al., 2015):

- Pheromone updating rule: Pheromone is a kind of variable used in ACO which is representing the Ants pheromone. Which they deposit on their path during food search when the next ant comes in contact with pheromone they know that they are on right track and add more pheromone for the next group and so on (Márquez et al., 2016).

This Algorithm updates the pheromone values after each iteration by using this Eq. 10 (Gao et al., 2015).

\[
\tau_{id}(t + 1) = (1 - \rho) \times \tau_{id}(t) + \Delta \tau_{id}
\]

where, \(\rho\) is evaporation rate; \(\tau_{id}(t + 1)\) is updated value; \(\tau_{id}(t)\) is last value of pheromone; \(\Delta \tau_{id}\) increase in pheromone.

- Test case selection rule: As per Ants’ behaviors, the movement of ants is irrelevant to all other ants. Similarly, an ACO algorithm selecting the next node is the most important step which uses pheromone quantity and then decided which node to move. So, the next test case selection is calculated using Eq. 11 (Gao et al., 2015).

\[
p_{id}^k(t) = \frac{[\tau_{id}(t)]^\alpha \times [\eta_{id}(t)]^\beta}{\sum_{k=1}^{n} [\tau_{id}(t)]^\alpha \times [\eta_{id}(t)]^\beta}
\]

110
where, $\eta_{dl}(t)$ is visibility function; $L_{j+1}^{k}$ is a list of alternative test cases. Where $\alpha$ and $\beta$ are the parameters to regulate the influence of pheromone and visibility respectively. Both ACOs techniques used really small programs with a set of 8-10 test cases having 10 faults only as shown in Table 2.

In TCP-ACA comparisons were made with particle swarm optimization (PSO) algorithm, genetic algorithm (GA), and random test methods which showed better results. On the other hand, Gao-ACO was only compared with random test case selection none of the other comparison algorithms were mentioned as shown in Table 3.

Following limitations were observed during the study of this algorithm. The first one is the type of construction used to validate the algorithm depends upon the results obtained. In order to validate the results, the APFDc matrix was used. Secondly, a new technique was not compared with any of the state-of-the-art methodologies and the third but most important one is that the technique is tested with a very small program with eight test cases having 10 faults only.

### 3.6. Modified ant colony optimization (m-ACO)

Solanki et al. (2015; 2016) proposed a modified ant colony optimization technique (m-ACO). In this algorithm when an ant finds food, it brings food back home and goes again to collect and this process repeats until all food sources are finished. Moreover, if the path is short all ants follow the same path until food finishes. This technique used the following test criteria code coverage, fault detected, and execution time. In order to achieve the maximum fault coverage, the APFD metric was used proposed by Elbaum et al. (2002). Algorithm proceeds as follows:

- Initialization of Ant variable
- Randomly search food source
- Calculate food fitness
- Calculate the probability of selection
- Calculate the reduction in pheromone

In order to validate m-ACO author used 3 small case studies were used. 1) 10 test cases with 10 faults 2) 9 test cases with 5 faults 3) 5 test cases with 5 faults. All the test case was executed on a Perl program as shown in Table 2. In m-ACO none of the comparisons were made with any state-of-the-art methodologies to verify the correctness of this algorithm. Instead of that only APFD was calculated with prioritized and un-prioritized test cases with showed comparatively better results than un-prioritized test cases as shown in Table 3.

This technique has the following limitations. First is only 3 small case studies with a limited number of test cases were used and it is highly recommended that this methodology should be validated with real-time and big data sets. By doing this we can get a real picture of results. The second is external validity. This approach is not compared with any of the state-of-the-art methodologies. It would be good to perform a comparison with some state-of-the-art prioritization methodologies. The third is the type of construction used to validate the algorithm depends upon the results obtained. In order to validate the results only the APFDc matrix was used, in which the APFD value of prioritized and un-prioritized results was compared.

### 3.7. ETS-based ACO algorithm for TCP

Epistasis-based Ant Colony Optimization technique is Multiobjective search-based regression Test Case Prioritization (MoTCP). This technique is a Metaheuristics that is inspired by nature (movement of ants to find food) (Gao et al., 2015). Study shows that epistasis theory also improved the effectiveness and efficiency in a single-objective GA (Deb et al., 2000). The use of epistasis theory in the multiobjective problem also showed promising results. ETS-Based ACO Algorithm for TCP is given below (Yuan et al., 2015).

#### Algorithm

Steps of ETS-Based ACO Algorithm for TCP.

- Setting of parameters of ACO
- Initialization of all the new sequences.
- Initialization of pheromone (weights of transitions).
- New test case sequence creation based on ant method equation 12.
- Test sequence evaluation using a fitness function.
- Ranking of test sequences as per NSGA-II method (Deb et al., 2000).
- Comparisons and Updation of the test sequence.
- Updating pheromone.

$$P_{ij}^k = \left\{ \begin{array}{ll}
\frac{[\tau_{ij}]^{\alpha}[\eta_{ij}]^{\beta}}{\sum_{l \in N} [\tau_{il}]^{\alpha}[\eta_{il}]^{\beta}} & \text{if } \tau_{ij} > 0 \\
0 & \text{otherwise}
\end{array} \right. \quad (12)$$

where, $\tau_{ij}=$pheromone (weight of the transitions); $d=$number of objectives; $\eta_{ij}=$heuristic information of an object; $\alpha=$heuristic factor; $\beta=$relative weight of heuristic value; $N[i]^k=$is the set of candidate test cases; $k=$ant.

In order to validate E-ACO author used the SIR program for unit test and the V8 program for the smoke test. The size of the V8 program is quite large which is good but it’s still not that large enough for big experiments as shown in Table 2.

E-ACO showed very impressive results as compared with the above-explained ACO algorithms. This new methodology was also compared with the state-of-the-art algorithms NSGA-II for MoTCP and ACO which shows the prominent results as shown in Table 3. E-ACO showed very prominent results still it has some limitations that must need to address to make it a standard algorithm. The first one is instrumentation accuracy, which could yield different results. It would be better if we test this using different software instead of using only gcov
(A GCC tool). Secondly, since the real-time program was not used in testing, in compensation to this test was run 10 times. It would be much better to use the real-time big program to get accurate analysis results. Third swarm sizes of different scales should be used. So we could easily compare the efficiency of ACO and E-ACO. The fourth and last one is the parameters configuration used in E-ACO which is the same used by NSGA-II. It would be rather good to change the parameters to verify the conclusion.

3.8. Hybrid ACO

This approach is proposed by Ahmad et al. (2018) which is the combination of two different approaches.

Step-I: In the first section by using test factors priorities are assigned to the test cases.

Step-II: Once we have priorities assigned to test cases we apply ACO to calculate the optimal sequence which has the lowest execution time but the highest fault rate.

The approach seems interesting but the study doesn't show anything about testing or if any comparisons were made with any state of art techniques.

3.9. Local beam search algorithm (LBS) for TCP

Jiang and Chan (2015) proposed a novel Local Beam Search Algorithm (LBS) which is an input-based search technique. LBS technique uses lightweight input as compared to other techniques like code-coverage. In each iteration of this algorithm width of the beam is kept constant that's the reason this approach is much more efficient than the other. For example, the greedy algorithm doesn't select the new test cases from uncovered test cases only directly increases cost and time (Elbaum et al., 2002). This input-based LBS algorithm takes randomly selected test case set as input, and processes it, and generates the set prioritized set of test cases.

This approach uses distance metric to calculate the distance between prioritized and un-prioritized test cases (Gusfield, 1997). The steps of the algorithm are given below (Jiang and Chan, 2015).

**Algorithm**

- Selection of candidate set which is not yet selected
- Calculation of distance matrix between the candidate and already prioritized ones
- Test cases are sorted in descending order.
- Appending the top test case in the successor pool.
- Repeating the process until all the test case has been prioritized.

LBS used 4 Software-artifact Infrastructure Repository (SIR) programs for testing. Three of them are UNIX-based programs (flex, grep, gzip, and sed) as shown in Table 2. Comparisons were made in LBS with state-of-the-art algorithms ART, Greedy, and Genetic Algorithm (GA). LBS showed more impressive results than GA and Greedy but less than the Adaptive Random Test case (ART) as shown in Table 3.

Results shown by LBS are very impressive but this algorithm has some limitations. 1) Platform used, during the testing of this approach only a C program was used. In order to cross-validate the finding other programming languages should be considered. 2) Testing Suites, during the analysis of this approach only branch-adequate test suites were used. The use of different test case suites could improve the results. 3) Use of distance metric, there are a lot of other metrics available like Hamming distance metric could also be tried to analyze results. 4) Since the LBS follows a randomized procedure so, the author repeats the testing 50 to get better results. It would be much more appropriate to use a big population of the test case to avoid such issues. 50 study shows that search-based TCP techniques don't use fault rate detections. While this research used APFD which make it difficult to access that whether the results produced by this research are consistent with previous studies or not.

3.10. Particle swarm optimization (PSO)

Particle Swarm Optimization (PSO) is a multi-object TCP technique which prioritizes test case based on changes in the code. PSO is an optimization methodology that was first introduced by Eberhart et al. (2001). Souza used the Binary Constrained PSO approach to prioritize test cases using the following three steps (Tyagi and Malhotra 2014; de Souza et al., 2010):

- Removal of extra test cases.
- Selection of test cases with the highest coverage but lowest execution time.
- Prioritizing test cases.

Hla et al. (2008) proposed an algorithm that performed test case prioritization based on changes in software sections by the PSO algorithm. This algorithm has been used a lot in various functional optimization problems because of its fast convergence property. This algorithm uses multi-objects to calculate fitness values using statement, branch, function, and code coverage as shown in Eq. 13. Also, it uses the velocity and position of the test cases for the next fitness evaluation which can be calculated using Eqs. 14 and 15. PSO algorithm has two parts (Hla et al., 2008).

**Part-I: Tokenizer**

- Take code as input
- Converts into token line by line
- Save into database

**Part-II: Prioritizer**

- Initialize priorities and fitness values
- Fitness value calculation using Eq. 13
- Position value calculation using Eq. 14, 15

\[ F(f) = E(f_a \wedge f_b \wedge f_c) \]  
(13)

where, \( f_a \) maximum statement coverage out of set of statements \( \{s_1, s_2, ..., s_n\} = \arg \max f_a(u_i, T, S, m) \); \( f_b \) total branch coverage out of set of compound conditions \( \{c_1, c_2, ..., c_n\} = \arg \max f_b(u_i, T, C, m) \); \( f_c \) total function coverage out of set of procedures \( \{p_1, p_2, ..., p_n\} = \arg \max f_c(u_i, T, P, m) \); \( f_d \) total code coverage out of set of software units \( \{u_1, u_2, ..., u_n\} = \arg \max f_d(u_i, T, U, m) \).

\[ V_{ik+1} = c_0 V_i + c_1 \text{rand.}(f_i - x_{ik}) + c_2 \text{rand.}(f_g - x_{ik}) \]  
(14)

\[ X_{ik+1} = X_{ik} + V_{ik+1} \]  
(15)

where \( x_i \) is the position vector and \( v_i \) is rate of change of position vector. \( c_0, c_1, \) and \( c_2 \) are weight factors.

PSO used 20 test cases from JUnit as shown in Table 2. A JUnit test suite provides automatic test case creation and execution as shown in Table 2.

Comparisons were made in PSO with state-of-the-art Greedy algorithm. PSO showed less run time complexity as shown in Table 3. Following limitations were observed during this study. 1) In order to keep internal validity in mind. This approach was tested using 20 test cases from JUnit which is really less. 2) Second is external validity, this new approach is only compared with the Greedy algorithm. It is highly recommended to validate the procedure using some real-time application and make a comparison with some other state-of-the-art techniques to get a better picture of results.

### 3.11. Weight Hybrid TCP using PSO (WH-PSO)

Khatsibyrbini et al. (2017) proposed this weight hybrid string distance technique. The main purpose of this technique was to get a higher APFD rate and efficiency in execution. A hybrid string distance was created using weight and distance between test cases. String distance is of two types: character-based and term-based. Two character-based string metrics used in this study are Manhattan Distance by Ledru et al. (2012) Eq. 16 and Levenshtein Distance by Jiang and Chan (2015) Eq. 17.

\[ \sum_{i=1}^{n} |x_i - y_i| \]  
(16)

\[ lev_{a,b}(i,j) = \begin{cases} \max(i,j) \\ \min(lev_{a,b}(i-1,j) + 1) \\ lev_{a,b}(i,j-1) + 1 \\ lev_{a,b}(i-1,j-1) + 1(a_i \neq b_i) \end{cases} \]  
(17)

Two term-based string metrics used are Cosine Similarity distance (Jiang and Chan, 2015) Eq. 18 and Jaccard Coefficient distance (Cha, 2007) Eq. 19.

\[ S_{\text{cos}} = \sum_{i=1}^{n} \frac{f_a(i) f_b(i)}{\sqrt{\sum_{i=1}^{n} f_a(i)^2 \sum_{i=1}^{n} f_b(i)^2}} \]  
(19)

Hybrid string distance is calculated by multiplying the distance with the next test case weight.

"Term frequency-inverse document frequency" (TFIDF) shows the importance of a word in the whole document.

\[ tf_{ik} = \frac{f_a}{\sum_{j=1}^{n} f_{jk}} \log(f_{ik}) \]  
(20)

\[ idf_{k} = \log \frac{n}{\text{no. of documents containing } k} \]  
(21)

The whole process is divided into three phases:

A. Information Extraction phase: In this phase test cases and their inputs are extracted. The extracted inputs are saved into a separate document which will be used so that they could be utilized in the next phase. After that, the test cases from both programs (original and version) are executed and a fault matrix sheet is prepared to compare both results.

B. String Distance Calculation: The extracted in the previous phase is used to calculate the string distances using Eqs. 16 and 17. The weight of test cases is also calculated using Eqs. 18 and 19 and the results of both calculations are populated into the test case distance matrix. Then hybrid string distance is calculated. After the completion of calculations, PSO is used to prioritize the distance matrix.

C. Evaluation phase: In this phase, a matrix sheet is created using results obtained in an earlier phase. WH-PSO used 3 Software-artifact Infrastructure Repository (SIR) programs for testing. Three of them are SIEMENs benchmark programs (tacas, jtcas, and cs-tacas) as shown in Table 2. This new approach was not compared with any state-of-the-art algorithms only comparison was made among character-based, term-based, weight-hybrid character-based, and weight-hybrid term-based results as shown in Table 3. Results shown by weight hybrid PSO are very impressive but it would be really good to test this approach with other states of the art methodologies.

### 3.12. Greedy particle swarm optimization (G-PSO)

Allawi et al. (2020) proposed this novel hybrid greedy and PSO algorithm (G-PSO) which is the combination of both greedy and PSO algorithms.

In this approach, the author creates the fitness function using Particle Swarm Optimization (PSO) and then applies the Greedy approach to find the partial best and makes them as global best and select the optimal solution from there. After this step, new algorithm moves back to the PSO approach where it finds the velocity of each particle is calculated using Eq. 22 and the location of the next partial using Eq. 23. This process repeats until the termination condition occurs.
\[ v_i(t) = w v_i(t) + c_1 r_1 [p b_i(t) - x_i(t)] + c_2 r_2 [g b_i(t) - x_i(t)] \]  
\[ x_i(t) = x_i(t) + v_i(t) \]  

Algorithm proceeds as follows (Allawi et al., 2020):

- Random Initialization of particles.
- Calculation of fitness function using PSO.
- Greedy part:
  - Determines and selects the particle with the partial Best position value to be placed as the global Best solution.
  - Selects the optimal solution from the available candidate solutions at a particular point.
  - Each particle velocity can be calculated using Eq. 22.
  - According to Eq. 23 move to the next position.
  - The algorithm will be stopped if the termination criteria are satisfied; otherwise, return to Step 2.

Six different Java card applications (JCS Applets) were used as data set to test the algorithm which includes a Network connection tracker, HelloWorld, PKI, RSA Crypto, Calculator, OATH, Passport, CoolKey as shown in Table 2.

The comparison was done between the GPSO algorithm and standard genetic algorithm (GA), which showed that GPSO exceeds the GA in various ways. GPSO takes a fewer number of iterations and time but gives us high coverage percentage as shown in Table 3.

Following are the limitations found during this study. 1) Accuracy of tools used to find the coverage information. The author used professional software in order to minimize the risk but it would be highly recommended to try to use some other software to just see the variety of results. 2) Most of the testing data used in this study is from the SIR repository. Only one program was used from ESA. 3) Results of this approach also depend upon the type of algorithm used for multi-object prioritization. The author tried to use so far best algorithm NSGA-II but it is highly recommended to test the approach with some other multi-object test case prioritization techniques like firefly etc. to test the variation in results.

### 3.13. Greedy algorithm (GA)

#### 3.13.1. Additional greedy algorithm (AGA)

Yoo and Harman (2007) proposed the Additional Greedy Algorithm. This technique uses NSGA-II and its variant which is called vNSGA-II algorithms.

The vNSGA-II has two main modifications as compared to NSGA-II which are as follows:

- vNSGA-II uses a separate group of sub-population which is good because different objectives can be assigned to each sub-group.
- vNSGA-II keeps the record of the best sub-population.

The additional greedy algorithm only supports up to two objects. In order to work with more than two objects, we have to combine the objects to make one by using the weight sum approach. That is the reason AGA doesn’t show good results for multi-object TCP.

AGA used 4 Software-artifact Infrastructure Repository (SIR) programs SIEMENs based programs (printtokens, printtokens2, schedule, schedule2), and 1 European Space Agency (ESA) program (space) as shown in Table 2. Comparisons were made in this study among two genetic algorithms, NSGA-II, and its variation. It has been seen that the greedy algorithm doesn’t work well for multi-objective TCP it is only good for single objects as shown in Table 3.

Following limitations were found during this study.1) Accuracy of tools used to find the coverage information. The author used professional software in order to minimize the risk but it would be highly recommended to try to use some other software to just see the variety of results. 2) Most of the testing data used in this study is from the SIR repository. Only one program was used from ESA. 3) Results of this approach also depend upon the type of algorithm used for multi-object prioritization. The author tried to use so far best algorithm NSGA-II but it is highly recommended to test the approach with some other multi-object test case prioritization techniques like firefly etc. to test the variation in results.
The proposed algorithm consists of two parts, graph generator, and graph traversal.

\[
\text{Node value} = \frac{\text{code coverage}}{\text{execution time}}
\]

Proposed two algorithms:

**Part-I: Graph Generation:**
1. Nodes creation.
3. Edge value calculation using Jaccard Distance Eq. 19.

**Part-II: Traversal Procedure:**
1. Select node with maximum value as a start point.
2. Calculate the highest gain using the traversal algorithm.
3. Repeat this process until all connection edges will eliminate.

In order to test this approach, the author used 4 open-source programs, among them 2 are obtained from SIR (jmeter, jtopas) and two of them are none SIR (nopCommerce, Umbraco-CMS) as shown in Table 2. Experiment results showed indicated approach is effective and efficient as compared with the unprioritized approach only as shown in Table 3.

There are few findings of this new methodology.
1. This new technique is not compared with any state-of-the-art TCP methodology to get a clear picture.
2. The choice of the distance function. The most popular method was used for it but it could give different results with a different function. It is highly recommended to test the new approach using other distance functions as well so, we could get a clear picture of this algorithm.

<table>
<thead>
<tr>
<th>S.No</th>
<th>Algorithms</th>
<th>Repository</th>
<th>Datasets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Firefly Algorithm</td>
<td>SIR</td>
<td>UNIX benchmarks Programs (flex, grep, gzip)</td>
</tr>
<tr>
<td>2</td>
<td>Optimal FA</td>
<td>Non-SIR</td>
<td>SIEIMENs benchmarks Programs (tcas, ctcas, j-cas)</td>
</tr>
<tr>
<td>3</td>
<td>Hybrid FA</td>
<td>SIR</td>
<td>4 small programs with set of 16-48 test cases.</td>
</tr>
<tr>
<td>4</td>
<td>Hypervolume Genetic Algorithm (HGA)</td>
<td>SIR</td>
<td>UNIX Benchmarks Programs (flex, grep, gzip)</td>
</tr>
<tr>
<td>5</td>
<td>ACO</td>
<td>Not SIR</td>
<td>A small program with set of 8-10 test cases with 10 faults only.</td>
</tr>
<tr>
<td>6</td>
<td>m-ACO</td>
<td>Not SIR</td>
<td>Perl Program based 3 case study used to test</td>
</tr>
<tr>
<td>7</td>
<td>ETS based ACO</td>
<td>SIR</td>
<td>3 UNIX Benchmarks Programs (flex, space, bash)</td>
</tr>
<tr>
<td>8</td>
<td>Hybrid ACO</td>
<td>None</td>
<td>Open-source JavaScript engine V8</td>
</tr>
<tr>
<td>9</td>
<td>LBS</td>
<td>SIR</td>
<td>4 UNIX Benchmarks Programs (flex, grep, gzip, sed)</td>
</tr>
<tr>
<td>10</td>
<td>PSO</td>
<td>Non-SIR</td>
<td>JUnit 20-test case</td>
</tr>
<tr>
<td>11</td>
<td>WH-PSO</td>
<td>SIR</td>
<td>Six Java Card Application Programs (Network connection tracker, HelloWorld, PKI, RSACrypto, Calculator, OATH, Passport, CoolKey)</td>
</tr>
<tr>
<td>12</td>
<td>Greedy PSO</td>
<td>Non-SIR</td>
<td>4 SIEIMENs Benchmarks Programs (printtokens, printtokens2, schedule, schedule2)</td>
</tr>
<tr>
<td>13</td>
<td>Additional Greedy</td>
<td>SIR</td>
<td>1 European Space Agency (space)</td>
</tr>
<tr>
<td>14</td>
<td>Enhanced Greedy</td>
<td>SIR</td>
<td>8 SIEIMENs Benchmarks Programs (Tcas, Totinfo, Replace, Schedule, Schedule2, Printtokens, Printtokens2, and Space)</td>
</tr>
<tr>
<td>15</td>
<td>graphite greedy</td>
<td>SIR, NON-SIR</td>
<td>4 UNIX Benchmarks Programs (jmeter, jtopas) nopCommerce, Umbraco-CMS</td>
</tr>
</tbody>
</table>

4. Research methodology

This study was divided into the following stages:
1. Research question.
2. Repositories Selection and Searching Strategies

4.1. Research questions

Following are the research questions of this paper:

**Q1:** How TCP is important in Software Testing?

**Q2:** What are the different classifications of test case prioritization techniques?

**Q3:** What are single object and multi-object TCP techniques?

**Q4:** What are the latest studies in TCP?

**Q5:** Which metrics are commonly used in TCP?

**Q6:** Which prioritization methods are used often and what are their proportions?

4.2. Repositories selection and searching strategies

In this study, we sued keywords like “Test Case Prioritization”, “Test Case Selection”, “Test Case Optimization” to search the well-known repositories like Google Scholar, IEEE Xplore, Elsevier, ACM Digital Library, and Springer as shown in Figs. 2, 3, 4, 5, 6, 7, and 8.
Table 3: Algorithms comparisons

<table>
<thead>
<tr>
<th>S.No</th>
<th>Algorithms</th>
<th>Compared With</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Firefly Algorithm</td>
<td>PSO, LBS, Greedy, GA</td>
<td>FF showed better results but LBS was almost close to FF.</td>
</tr>
<tr>
<td>2</td>
<td>Optimal FA</td>
<td>ACO</td>
<td>FF Algorithm showed better results.</td>
</tr>
<tr>
<td>3</td>
<td>Hybrid FA</td>
<td>FA, PSO, and Greedy</td>
<td>HGA showed prominent results over FA and PSO.</td>
</tr>
<tr>
<td>4</td>
<td>Hypevolume GA</td>
<td>Additional Greedy Algo, GA(AUC metric), NSGA-II, GDE3, MOEA/D-DE</td>
<td>HGA is more cost-effective and efficient in dealing with more criteria. In TCP-ACA comparisons were made with particle swarm optimization algorithm, genetic algorithm, and random test methods which showed better results. On the other hand, Gao-ACO was only compared with random test case selection none of the other comparison algorithms were mentioned. Prioritized with m-ACO and un-prioritized showed better results.</td>
</tr>
<tr>
<td>5</td>
<td>ACO</td>
<td>PSO and Random Test Case Selection</td>
<td>E-ACO algorithm has significant improvement effectiveness and the efficiency to the state-of-the-art NSGA-II algorithm. better results as compared to ACO LBS showed more efficient results than GA and Greedy but less than ART PSO has time complexity less than Greedy Algorithm Hybrid PSO showed impressive results</td>
</tr>
<tr>
<td>6</td>
<td>m-ACO</td>
<td>ACO</td>
<td>E-ACO algorithm has significant improvement effectiveness and the efficiency to the state-of-the-art NSGA-II algorithm. better results as compared to ACO LBS showed more efficient results than GA and Greedy but less than ART PSO has time complexity less than Greedy Algorithm Hybrid PSO showed impressive results</td>
</tr>
<tr>
<td>7</td>
<td>ETS-ACO</td>
<td>Original ACO for MoTCP</td>
<td>E-ACO algorithm has significant improvement effectiveness and the efficiency to the state-of-the-art NSGA-II algorithm. better results as compared to ACO LBS showed more efficient results than GA and Greedy but less than ART PSO has time complexity less than Greedy Algorithm Hybrid PSO showed impressive results</td>
</tr>
<tr>
<td>8</td>
<td>Hybrid ACO</td>
<td>NSGA-II algorithm for MoTCP</td>
<td>E-ACO algorithm has significant improvement effectiveness and the efficiency to the state-of-the-art NSGA-II algorithm. better results as compared to ACO LBS showed more efficient results than GA and Greedy but less than ART PSO has time complexity less than Greedy Algorithm Hybrid PSO showed impressive results</td>
</tr>
<tr>
<td>9</td>
<td>LBS</td>
<td>Genetic, Greedy(Total and Additional), Optimal, Hill Climbing and ART</td>
<td>LBS showed more efficient results than GA and Greedy but less than ART PSO has time complexity less than Greedy Algorithm Hybrid PSO showed impressive results</td>
</tr>
<tr>
<td>10</td>
<td>PSO</td>
<td>Greedy Algorithm</td>
<td>LBS showed more efficient results than GA and Greedy but less than ART PSO has time complexity less than Greedy Algorithm Hybrid PSO showed impressive results</td>
</tr>
<tr>
<td>11</td>
<td>WH-PSO</td>
<td>Genetic Algorithm</td>
<td>GP50 outperform the GA in terms of the average number of iterations, execution time, and coverage percentage Greedy algorithm is not good for multi objective TCP EAGA showed better results as compared to AGA but only for a single objective. High Average fault detection results</td>
</tr>
<tr>
<td>12</td>
<td>Greedy PSO</td>
<td>Genetic Algorithm</td>
<td>GP50 outperform the GA in terms of the average number of iterations, execution time, and coverage percentage Greedy algorithm is not good for multi objective TCP EAGA showed better results as compared to AGA but only for a single objective. High Average fault detection results</td>
</tr>
<tr>
<td>13</td>
<td>Additional Greedy</td>
<td>NSGA-II</td>
<td>Greedy algorithm is not good for multi objective TCP EAGA showed better results as compared to AGA but only for a single objective. High Average fault detection results</td>
</tr>
<tr>
<td>14</td>
<td>Enhanced Greedy</td>
<td>AGA</td>
<td>Greedy algorithm is not good for multi objective TCP EAGA showed better results as compared to AGA but only for a single objective. High Average fault detection results</td>
</tr>
<tr>
<td>15</td>
<td>Graphite</td>
<td>None</td>
<td>Greedy algorithm is not good for multi objective TCP EAGA showed better results as compared to AGA but only for a single objective. High Average fault detection results</td>
</tr>
</tbody>
</table>

Fig. 2: Search results of google scholar (TCP, TCS, TCO)
Fig. 3: Total search results of google scholar
Fig. 4: Search results of IEEE (TCP, TCS, TCO)
Fig. 5: Total search results of IEEE
5. Results

In this section each research questions were answered:

**Q1:** How TCP is important in Software Testing?
In Regression Testing (RT), the TCP is the most difficult phase of the RT that allows testers to risk management, test planning, cost-value analysis of the testing phase. Different approaches are used to prioritized test cases for example time, cost, effort, code coverage, and fault coverage which increase RT performance. By using these techniques test cases with the highest priority are selected for RT suites that give us the highest rate of fault detection.

**Q2:** What are different classifications of test case prioritization techniques?
Fig. 2, shows the detailed classification of the test case prioritization techniques.

**Q3:** What are single object and multi-object TCP techniques?
Table 4 shows which techniques are multi-object TCP and single-object TCP. This table also shows which metrics were used to validate the algorithm.

**Q4:** What are the latest studies in TCP?
Table 5 shows the latest studies in Test Case Prioritization techniques.

**Q5:** Which metrics are commonly used for TCP?
About 70% of the multi-object test case prioritization techniques used the Average Percentage Fault Detection (APFD) metric as shown in Table 4.

**Q6:** Which prioritization methods are used often and what are their proportions?
Table 6, shows the percentage of each algorithm from different repositories.

<table>
<thead>
<tr>
<th>Techniques</th>
<th>Algorithms</th>
<th>Metrics Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal Firefly Algorithm</td>
<td>Firefly Algorithm</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>Hybrid FA</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>Hypervolume GA</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>ACO</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>m-ACO</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>ETS based ACO</td>
<td>APFD, EET</td>
</tr>
<tr>
<td></td>
<td>Hybrid ACO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>LBS</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>PSO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>WH-PSO</td>
<td>APFD</td>
</tr>
<tr>
<td></td>
<td>Greedy PSO</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Additional Greedy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enhanced Greedy</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Graphite greedy</td>
<td></td>
</tr>
</tbody>
</table>

Table 4: Metrics used in different algorithms
Table 5: Latest test case prioritization techniques

<table>
<thead>
<tr>
<th>S.No</th>
<th>Ref ID</th>
<th>Algorithm</th>
<th>Year</th>
<th>Repository</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Shah et al.</td>
<td>Optimal Firefly Algorithm</td>
<td>2016</td>
<td>Elsevier</td>
</tr>
<tr>
<td>2</td>
<td>Khatibsyarbini et al.</td>
<td>Firefly Algorithm</td>
<td>2019</td>
<td>IEEE</td>
</tr>
<tr>
<td>3</td>
<td>Su et al.</td>
<td>Hybrid Firefly Algorithm</td>
<td>2020</td>
<td>IEEE</td>
</tr>
<tr>
<td>4</td>
<td>Di Nucci et al.</td>
<td>Hypervolume GA</td>
<td>2018</td>
<td>IEEE</td>
</tr>
<tr>
<td>5</td>
<td>Zhang et al.</td>
<td>Ant Colony Optimization</td>
<td>2019</td>
<td>IEEE</td>
</tr>
<tr>
<td>6</td>
<td>Solanki et al.</td>
<td>Modified ACO</td>
<td>2016</td>
<td>Springer</td>
</tr>
<tr>
<td>7</td>
<td>Bian et al.</td>
<td>ETS Based ACO</td>
<td>2017</td>
<td>IEEE</td>
</tr>
<tr>
<td>8</td>
<td>Ahmad et al.</td>
<td>Hybrid ACO</td>
<td>2018</td>
<td>Elsevier</td>
</tr>
<tr>
<td>9</td>
<td>Jiang and Chan</td>
<td>LBS</td>
<td>2015</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Tyagi and Malhotra</td>
<td>PSO</td>
<td>2014</td>
<td>ICSPCT</td>
</tr>
<tr>
<td>11</td>
<td>Khatibsyarbini et al.</td>
<td>Hybrid PSO</td>
<td>2017</td>
<td>JATIT</td>
</tr>
<tr>
<td>12</td>
<td>Allawi et al.</td>
<td>G-PSO</td>
<td>2020</td>
<td>Springer</td>
</tr>
<tr>
<td>13</td>
<td>Yoo and Harman</td>
<td>Additional Greedy Algorithm</td>
<td>2007</td>
<td>ACM</td>
</tr>
<tr>
<td>14</td>
<td>Hsu et al.</td>
<td>Severity-weighted Greedy Algorithm</td>
<td>2014</td>
<td>IEEE</td>
</tr>
<tr>
<td>15</td>
<td>Azizi and Do</td>
<td>Graphite Greedy Algorithm</td>
<td>2018</td>
<td>Elsevier</td>
</tr>
</tbody>
</table>

Table 6: Algorithm wise TCP percentage

<table>
<thead>
<tr>
<th>S.No</th>
<th>Algorithm</th>
<th>Google Scholar</th>
<th>IEEE</th>
<th>Springer</th>
<th>Total</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Firefly</td>
<td>15800</td>
<td>30</td>
<td>867</td>
<td>16697</td>
<td>4.5</td>
</tr>
<tr>
<td>2</td>
<td>GA</td>
<td>238000</td>
<td>1105</td>
<td>21076</td>
<td>260181</td>
<td>7.05</td>
</tr>
<tr>
<td>3</td>
<td>ACO</td>
<td>16900</td>
<td>111</td>
<td>1998</td>
<td>19009</td>
<td>5.1</td>
</tr>
<tr>
<td>4</td>
<td>LBS</td>
<td>19100</td>
<td>20</td>
<td>635</td>
<td>19755</td>
<td>5.3</td>
</tr>
<tr>
<td>5</td>
<td>PSO</td>
<td>16700</td>
<td>556</td>
<td>6396</td>
<td>23652</td>
<td>6.4</td>
</tr>
<tr>
<td>6</td>
<td>Greedy</td>
<td>16800</td>
<td>193</td>
<td>12468</td>
<td>29461</td>
<td>7.9</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>323300</td>
<td>2015</td>
<td>43440</td>
<td>368755</td>
<td></td>
</tr>
</tbody>
</table>

Table 7 shows the inclusions and exclusions criteria of this study.

Table 7: Inclusions and exclusions criteria of this study

<table>
<thead>
<tr>
<th>Inclusion criteria</th>
<th>Exclusion Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Regression Testing and/or Test case) and (Prioritization or Selection) and &quot;Name of Algorithm&quot; and (Journals or Conference Paper) and (IEEE or Springer or Elsevier)</td>
<td>Non-Software Testing and/or Non-English paper</td>
</tr>
</tbody>
</table>

5.1. State of the art challenges

5.1.1. Measurement

- Effectiveness: Most of the TCP techniques are validated in terms of their effectiveness by using APFD metrics. However, APFD is not quite suitable and has quite constraints. For example, in APFD metrics all test cases and faults have the same cost (Askarunisa et al., 2010). On the other hand, APFDc works well because it allows variable test cases and fault costs. But it only allows two parameters. It is recommended to use such metrics which allow multiple parameters. There is another metric that could be utilized like Average Percentage Statement Coverage (APSC), Average Percentage Branch Coverage (APBC), Average Percentage Loop Coverage (APLC), and Average Percentage Condition Coverage (APCC) (Li et al., 2007).

- Efficiency: Efficiency is also an important issue and there is no such proper metric use to measure the efficiency of a technique. So far only things considered are actual prioritization time and test-case execution time. It is not enough to consider these two parameters because most techniques require more information e.g., code coverage, etc. which may add more cost. Matrix-like ANOVA and ANCOVA (Nayak et al., 2019) and Average Percentage of Combinatorial Coverage (APCC) which considers combination weights and test costs (Wang et al., 2011). It is highly recommended to keep these things in mind before start prioritization so we could get accurate cost and computation time.

5.1.2. Data utilization

Study shows that effectiveness and efficiency of a logarithm depend upon various factors for example programming language and its framework, coverage, types of faults, etc. (Hao et al., 2016). Other than these factors the most important thing that could affect the effectiveness and efficiency of an algorithm is as follows:

- Size of data set: Most of the above describe studies are only using a small set of SIR programs like bash, flex, grep, gzip, sed, etc. and some of them even used very small programs having 8-10 test cases (Khatibsyarbini et al., 2019; 2017; Su et al., 2020; Di Nucci et al., 2018; Jiang and Chan, 2015; Yoo and Harman, 2007).

- Incomplete data: Second most important thing is input data used in the algorithm. For example, the results of the APFD matrix depend upon the distance and weight of test cases which depends upon input given to the system (Khatibsyarbini et al., 2019). Test cases have a lot of other data available in them. But it is very difficult, time-consuming, and expensive to utilize all the available data. The best example is system coding. The extraction and utilization of complete data could be a big challenge in the test case prioritization techniques which could give us different results.

- Parameter Tuning: The tuning of parameters is also an important issue and there is no such proper matrix use to measure the efficiency of a technique. But it is highly recommended to keep these things in mind before start prioritization so we could get accurate cost and computation time.
is as in the previous work (Li et al., 2007). Tuning to that parameter could influence the overall results. Lu studies the influence of change in code or its parameter. Study shows that change in code and parameter could influence the effectiveness an algorithm (Lu et al., 2016).

5.1.3. Execution time

The study shows that most of the techniques are tested using the SIR program which is a small set of programs. These programs have test cases whose execution time is very low. It would be worth testing the above-given prioritization techniques on a test case where execution time is high. But testing of high execution time test cases could increase the cost. It would be better to consider the individual cost of test cases instead of the total cost.

6. Conclusion and future work

In this paper, an empirical survey has been conducted on Test case prioritization techniques. In this study, we review 40 TCP papers published in journals and conferences with a systematic process. The main objective of this study is to explore and determine which approaches have been studied in detail and which of them are left behind and where is more chance of improvement. We divided this study into the following stages. In Stage-I we created six research questions. Stage-II we selected repositories from where we are going to get our literature. Stage-III we created a search methodology in which we search the most recent papers on state-of-the-art techniques. Stage-IV we setup inclusion and exclusion criteria which we utilized to search our data as shown above. The following suggestion was provided in this study:

- **Latest public data set utilization**: Almost 60% of the papers selected in this study are using a public data set which is good but data is being updated regularly which is not satisfactory. By the development of different programming languages and programming techniques updated code must be utilized for testing. It looks like becoming a trend of using SIR data instead of actual data or updated public datasets. Although there are challenges of getting actual data set and re-testing by other researchers.

- **Consideration of multi-object TCP approaches**: Study shows that 35% of the approaches are well-defined multi-object and 20% of them are purely single object TCP techniques and the rest of them are not properly stated as single or multi objects. It would be highly recommended to create a multi-object model-based approach that could set standards.

- **Consideration of cost criteria**: Survey showed that that most of the studies focus on code base techniques and less focus is given to cost criteria. Some studies used cost criteria as a total cost of execution of test cases. It would be better to consider the individual cost of test cases instead of the total cost.

- **State of the Art Techniques Comparison**: Study showed that 46% of the techniques have only performed a comparison with the well-defined state-of-the-art methodologies. 26% of the approaches have not compared any state-of-the-art technique. It would be highly recommended to set up such models which accept the studies having compared with the state-of-the-art techniques. Also, it is good to conduct more studies for comparison of TCP techniques.

- **Survey study for Multi-Object TCP**: Study showed that none of the review/survey studies were conducted on multi-object TCP techniques. It would be highly recommended to separate publishing the review study of multi-object TCP techniques.

- **Utilization of real-time dataset**: Study showed that none of the approaches has been tested using any industrial or real-time dataset. It would be highly recommended to validate the test case prioritization techniques using some big industrial or real-time data set to get a clear picture of the new approach. However, only 10% of the studies used non-SIR data set but not industrial data set.

**Compliance with ethical standards**

**Conflict of interest**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**References**


PMid:20799183

PMcid:PMC6469475