Nature Inspired Approaches for Test Case Selection in Regression Testing: A Review

Sonam Chaudhary
Department of Computer Science & Engineering
Sharda University
Uttar Pradesh, India
2021372682.sonam@pg.sharda.ac.in

Ankur Choudhary
Department of Computer Science & Engineering
Sharda University
Uttar Pradesh, India
ankur.choudhary@sharda.ac.in

Jyotsna Seth
Department of Computer Science & Engineering
Sharda University
Uttar Pradesh, India
jyotsna.seth@sharda.ac.in

Abstract— **To ensure that the software satisfies all necessary criteria, software testing is done. Regression Testing is an important part of software testing as it plays a crucial role in software testing procedure. It's done to make sure that everything works as it should, such that test cases with greater potential to discover faults are given higher priority. Test case selection is introduced to choose the most effective subset of test cases from a test suite, test case selection reduces the total cost, time, and effort necessary in the process of software testing. It works by removing the obsolete and redundant test cases. Within all the computational intelligence algorithms the nature inspired algorithms are highly demanded due to their increased efficiency in solving complex problems. Nature Inspired Approaches refers to the approaches derived directly from the nature. Nature acts as the main branch of knowledge that provides various solution for the complex problems. Incorporation of Nature Inspired Approaches in test case selection problem generates a subset of test cases from the test suite resulting in time, cost and effort reduction. With the use of research questions, this paper offers a thorough review of Test Case Selection methods utilizing Nature Inspired Approaches.**

Keywords—Software Testing; Regression Testing; Test Case Selection; Nature Inspired Approaches

I. INTRODUCTION

Software testing is crucial to the Quality Assurance process that creates error-free software. To remain competitive in today's world and to meet the ever-changing needs of clients, software updates are unavoidable, as are their adverse effects. The most important need from now on is to examine whether or not the alterations have introduced any aspect repercussions. It is almost hard to do millions of executions each day on bigger applications [1]. Due to the need for human resources for setup, execution and monitoring of test cases testing becomes expensive. With each update, the software becomes increasingly complex which is difficult to check exhaustively because of restrained time and different resources. The improvisation of regression testing has been carried out where the researchers had developed various techniques to improve the terms of regression testing. These techniques are discussed and formally described by Yoo and Harman (2012) [2]:

Test suite Minimization/Reduction (TSM): These Techniques aims to minimize the size of the test suite by elimination of redundant test cases. However, a decrease in the number of test cases may result in a decline in the capacity to identify error.

Test Case Prioritization (TCP): These techniques aim at identification of an effective order of test cases in such a way that the test cases with a greater capacity for defect detection are given a higher priority.

Test Case Selection (TCS): This technique works by selecting a subset of valid test cases from the initial test suite based on the given criterion. It complies with all standards while reducing the number of test cases that must be run. By eliminating the unnecessary and duplicate test cases, it is accomplished.

While both test suite minimization and test case prioritization seek to cut down on testing time and expense, they might also omit certain important test cases that can find errors and so contribute to low efficiency. But with great efficiency and effectiveness, test case selection algorithms choose a subset of test cases from the test suite. The regression test selection problem has been clearly characterized by Rothermel and Harrold as: Let P be an application program and P′ be a modified version of P. Let T be the test suite developed initially for testing P. An RTS technique aims to select a subset of test cases T′⊆T to be executed on P′, such that every error detected when P′ is executed with T is also detected when P′ is executed with T′ [3].

Nature is a bulk producer of knowledge and concepts. Computational Intelligence algorithms are faming amongst traditional optimization. The researchers are trying to solve the complex problems with natural phenomenon which is giving a positive outcome. Nature-inspired algorithms are those that mathematically express the issue by replicating natural events and deliver the best answers [4]. Various Nature Inspired techniques are being used in the sector right now. There are various inspiration from nature like biology, Physics, chemistry and social Phenomenon's. The common denominator for all NP-hard problems is that they cannot be solved in an acceptable time period using typical deterministic methodologies. They necessitate a different, stochastic strategy if the task is to be addressed in an acceptable amount of time. The group of stochastic algorithms includes metaheuristic approaches. A metaheuristic is an advanced method or heuristic that can solve an optimization issue sufficiently well. They typically

function by sampling a subset of solutions that are too numerous to be fully enumerated. A well-known and significant family of metaheuristics methods are nature-inspired algorithms. Numerous real-world NP-hard jobs have been handled using metaheuristic methods. Nature-inspired metaheuristics can be divided into two categories: Evolutionary algorithms and the swarm intelligence (SI) metaheuristics. Nature Inspired Metaheuristics have always been proved to be game changers. The motivation of EA approaches come from Natural selection and the hypothesis that only the most fit members in a group would endure. SI metaheuristics, the last family of nature-inspired techniques, were inspired by the behavior of groups of very basic creatures. When vast groups of these small animals assemble, they may self-organize, demonstrate great levels of coordination, and perform complicated tasks (swarms). This property of swarming is the major motivation for all of the SI algorithms [5].

This paper's goal is a comprehensive analysis of several Nature Inspired Approaches used for Test Case Selection in regression testing. It examines the most investigated regression testing methods, the most used platform. Also, it provides a study for the nature inspired metaheuristic algorithm that is highly investigated, most utilized dataset and evaluate the basis for improvement of test case selection methods using metaheuristics. As a consequence, we carried out a systematic mapping research, a technique used to identify, choose, synthesize, summarize, and assess all pertinent problems and solutions connected to test case selection [2]. The Four Research Questions that structure this survey are stated below:

RQ1.Most Investigated Regression Testing Method.

RQ2. The Most frequently used Dataset.

RQ3. The Most frequently applied Nature Inspired Metaheuristic algorithm.

RQ4. Test Case Selection using nature inspired approach papers are mostly Published in journal or conference.

II. LITERATURE SURVEY

Software Testing serves as the crucial activity in software cycle. Also, it's a never-ending process and continue to become complex and expensive after every update. Testing becomes a tedious task because a tester has to test the updates along with the previous functionalities which is in a way very time consuming.

With each regression cycle more test cases are added and then it becomes complicated especially in the case of larger software. To overcome this problem regression testing techniques are taken into scenario. While attempting to cut costs and time, the reduction and prioritization of some of the test case can result into low fault detection capability. So, Test Case Selection stands out in this criterion as it works by selecting a subset of test case by keeping into validation the efficiency of the software. The conventional methods do not prove to be very effective and are very time consuming. So, Nature Inspired approaches are used for this problem instead of the conventional approaches.

This paper systematically reviews Test Case Selection techniques using Nature Inspired Approaches to evaluate the progress of the techniques which are currently used in the situation and what different is needed from the state of art in research. This paper provided four research questions, and with the help of these four questions the review is carried out.

Table 1: Literature Survey								
Ref. No	**Author**	**Year**	**Approach**	**Data set**	**Metaheuristic Approach**	**Findings of the study**	**Single or Multi objective**	**Conference / Journal**
1	Yoo et al.[6]	2007	TCS	SIR	Single objective GA, NSGA-II, vNSGA-II	This paper gives an introduction of the concept of Pareto Efficiency for test case selection.	Multi-objective	Journal
2	Cartaxo et al.[7]	2007	TCS	LTS	Works using a similarity function	The introduction of an automatic strategy to test case selection which is based on similarity between the test cases is presented in this paper.	Single-objective.	Journal
3	Maia et al.[8]	2009	TCS	SIR	NSGA-II	This paper gives an proposal to an multi-objective formulation for the regression test case selection problem by using multi-objective genetic algorithm and then compared it with a random algorithm.	Multi-objective	Conference

4	Biswas et al.[9]	2010	TCS	NA	NA	This Paper reviews the techniques for various categorized programs.	Multi-Objective	Journal
5	Mirarab et al.[10]	2012	TCS	SIR (5 Java Programs)	Greedy algorithm	This paper introduces the approach to select and order predetermined number of test cases from a test suite and introduce 3 novel techniques.	Multi-objective	Conference
6	Beheshti et al.[11]	2013	TCS	NA	Metaheuristic Approach	This paper provides an overview and comparison of Meta-Heuristics algorithms.	NA	Journal
7	Augustsson et al.[12]	2013	TCS	TOY	Conventional Methods	This paper gives an introduction and comparison to the framework which is supposed to evaluate regression test selection techniques	Multi-objective	Journal
8	Malhotra et al.[13]	2014	TCS, TCP, TSM	SIR	Mutated Artificial Bee Colony	This essay provides a succinct comparison of various Meta-Heuristic Algorithms and proposed MABC Algorithm.	Multi-Objective	Journal
9	Panichella et al.[14]	2015	TCS	SIR	DIV-GA, NSGA-II	This paper proposes a novel diversity- preserving technique.	Multi - objective	Conference
10	Lawanna et al.[15]	2016	TCS	SIR	Proposed a model using conventional methods	This paper gives an algorithm which chooses the test cases with high efficiency and a comparison with other techniques is done.	Multi-objective	Conference
11	Engström et al.[16]	2016	TCS	NA	NA	This paper gives a brief review for regression test selection techniques given in different papers.	NA	Conference
12	Kazmi et al.[17]	2017	TCS	NA	NA	This paper presents literature review of the existing research in regression test case selection techniques by examining the empirical studies.	NA	Journal
13	Choudhary et al.[18]	2018	TCS	SIR	Pareto based Multi-objective harmony Search	This paper proposes a regression test case selection strategy using Pareto based multi-objective Harmony Search Algorithm.	Multi-objective	Conference
14	Yadav et al.[19]	2019	TCS, TCP	TOY	GA	This paper presents a technique for regression testing using UML diagrams and the code-based analysis for object-oriented software.	Multi-objective	Conference

15	Gokilavani et al.[20]	2021	TCS	SIR	RBCH, DBT	This paper presents a RBCH algorithm to select Ch and DBT Algorithm to prioritise the test case clusters.	Multi-objective	Journal
16	Mangave et al.[21]	2021	TCS	SIR	Greedy Algorithm	This paper proposes an approach for regression testing based on two criteria.	Multi-Objective	Journal
17	Anu Bajaj et al.[22]	2019	TCS, TCP, TSM	NA	NA	This paper explores the work contributed by researchers on regression testing using nature inspired approaches.	NA	Conference
18	Catal et al.[2]	2012	TCP	NA	NA	This paper provides a wide review on the kind of techniques that have been used for test case prioritization and determine the area for scope of improvement.	NA	Journal
19	Singhal et al.[23]	2021	TCS, TCP	NA	NA	This review presents a systematic literature of the methods and techniques which are utilized in regression test cases selection and prioritization.	NA	Journal
20	Nagar et al.[24]	2014	TCS, TCP	TOY	Hybrid PSO, GA	This paper introduced novel hybrid PSO for selection and prioritization using PSO AND GA.	Multi-objective	Conference

III. FINDINGS AND DISCUSSION

This phase of research provides the answers to the four research questions with context to the information from the table and the questions are answered with the help of diagram for better understanding and interpretation.

RQ1. Most Investigated Regression Testing Method.
This research question's goal was to catalogue and classify accessible regression testing methods. This question provides a clear idea on the methods of Regression testing methods. Among the available regression testing methods, it was found from the study that Test Case Selection is the most investigated method. This is due to efficiency of test case selection method for the underlying problem of unnecessary and duplicate test cases which makes it difficult to test especially in case of the larger software's. To address the root issue Test Case Selection removes the obsolete and redundant test cases by selecting a valid subset of test cases from entire test suite with accordance to requirements and the specifications. This method simply aims to decrease the number of test cases that must be run. This method is in great

demand. for its effective and efficient results. By removing unnecessary and duplicate test cases, this approach selects a subset of test cases that is equally as effective as the original test suite. It does not prioritize or reduce the test cases just like the other methods. The researchers are investigating this method due to its efficiency in reducing the time, cost and efforts required for testing.
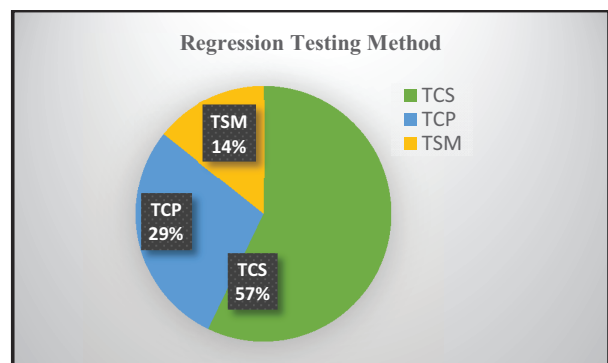

Fig 1: Regression Testing Methods

RQ2. The most frequently used Dataset.

It is well known that software engineering experiments must be reproducible and verifiable, hence the best datasets for evaluation are those that are available to the general public. Without this, it is impossible to access the data and determine if the experiment can be repeated. According to the study from papers or articles it is verifiable that SIR is the most utilized dataset as compared to any other.
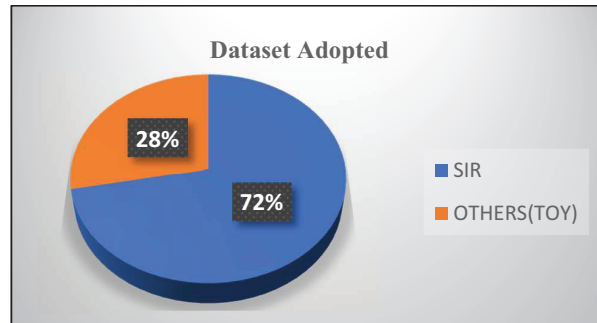

Fig 2: Dataset Adopted

RQ3. The Most frequently applied Nature Inspired Metaheuristic algorithm.

The number of test cases will rise in each regression cycle, increasing the size of the test suite as a result. Because if we add or update any new functionality into the software it needs to be tested and along with this the previous functionalities also needs to tested to guarantee the software's dependability and effectiveness. To solve this problem, we will make use of Test Case Selection method. But the data states that the conventional methods does not perform well with the larger suite size. To overcome this Nature inspired Approaches are taken into account. Metaheuristics are higher level procedure that aims to provide comparably good solution to the optimization problem. Metaheuristics works in a manner that it finds good solutions with minimal computational efforts with a technique to search over larger set of the feasible outcomes. Because of its unbelievable results over the certain years the researchers are using Nature inspired Metaheuristic Approaches to mend Test Case Selection problem in Regression Testing. This research provides an answer to which of the existing Nature inspired Metaheuristic Approach is the most investigated as compared to the other metaheuristic approach. NSGA-II is the algorithm which is highlighted in the research work carried out by the researchers and stands out in comparison to the other algorithms.
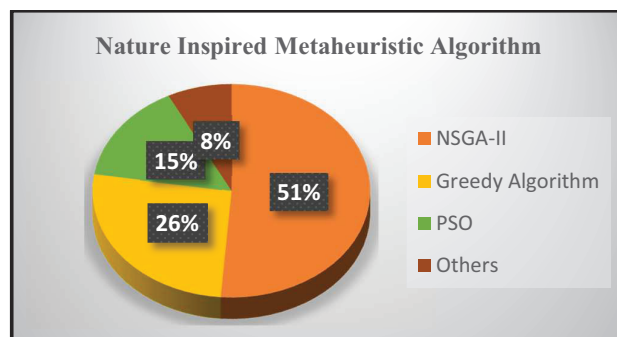
RQ4. Test Case Selection using nature inspired approach papers are mostly Published in journal or conference.

This research question helps us to understand that maximum number of publications are in journal or conference. This will help the future researchers to easily go through the content related to the Nature inspired approach and test case selection. To answer this accruing to the data collected the most of the papers are published in conference. Conference hold the maximum number of publications of the paper or article related to this topic.
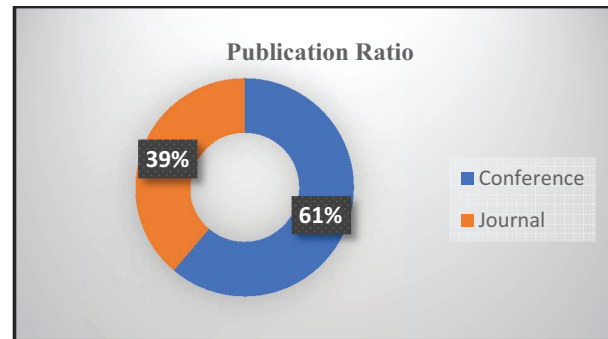

Fig 4: Publication Ratio

IV. CONCLUSION

The software keeps growing and the updates continue to evolve with every passing day and with this it becomes exhaustively complicated to test the newly added functionalities and retest the previous ones. To address this problem the optimization algorithms can be evaluated. In this survey we searched for the studies that focus on solving Test Case Selection problem using Nature Inspired Approaches. The conducted study was segregated into four research questions. We found that Test Case Selection serves the most investigated regression testing method amongst the others as it centres on selecting a subset of test cases form the test suite as per criteria while maintaining efficiency of the software. SIR Dataset serves to be the most utilized dataset among the researchers. Also, NSGA-II is the most frequently used Nature Inspired Metaheuristics Algorithm which proves to outstand other algorithms. Most of the studies related to Test Case Selection are published in conference. As a conclusion Nature inspired algorithms can solve Regression Test Case Selection problem with high accuracy. There is still more potential for further study and research on Metaheuristic Nature Inspired Approaches. The future works will direct towards the more extensive study about New Metaheuristic Nature Inspired algorithms.

Fig 3: Nature Inspired Metaheuristic Algorithm

REFERENCES

[1] A. Bajaj and O. P. Sangwan, "Nature-Inspired Approaches to Test Suite Minimization for Regression Testing," in *Computational Intelligence Techniques and Their Applications to Software Engineering Problems*, CRC Press, 2020, pp. 99–110. doi: 10.1201/9781003079996-7.

[2] C. Catal and D. Mishra, "Test case prioritization: A systematic mapping study," *Software Quality Journal*, vol. 21, no. 3, pp. 445–478, Sep. 2013, doi: 10.1007/s11219-012-9181-z.

[3] G. Rothermel and M. J. Harrold, "A Framework for Evaluating Regression Test Select ion Techniques*."

[4] I. Fister, X.-S. Yang, I. Fister, J. Brest, and D. Fister, "A Brief Review of Nature-Inspired Algorithms for Optimization," Jul. 2013, [Online]. Available: http://arxiv.org/abs/1307.4186

[5] M. Žižović, T. Živković, N. Bačanin Džakula, and I. Štrumberger, "Nature-Inspired Approaches in Software Testing Optimization," Jul. 2021, pp. 28–33. doi: 10.15308/sinteza-2021-28-33.

[6] D. S. Rosenblum, ACM Digital Library., and ACM Special Interest Group on Software Engineering., *Proceedings of the 2007 international symposium on Software testing and analysis.* ACM, 2007.

[7] E. G. Cartaxo, F. G. O. Neto, and P. D. L. Machado, "Automated Test Case Selection Based on aSimilarity Function."

[8] C. Loiola, B. Maia, R. A. Ferreira Do Carmo, F. Gomes De Freitas, G. A. Lima De Campos, and J. Teixeira De Souza, "A MULTI-OBJECTIVE APPROACH FOR THE REGRESSION TEST CASE SELECTION PROBLEM."

[9] S. Biswas, R. Mall, M. Satpathy, and S. Sukumaran, "Regression Test Selection Techniques: A Survey," 2011.

[10] S. Mirarab, S. Akhlaghi, and L. Tahvildari, "Size-constrained regression test case selection using multicriteria optimization," *IEEE Transactions on Software Engineering*, vol. 38, no. 4, pp. 936–956, 2012, doi: 10.1109/TSE.2011.56.

[11] Z. Beheshti, S. Mariyam, and H. Shamsuddin, "A Review of Population-based Meta-Heuristic Algorithms," 2013. [Online]. Available: www.i-csrs.org

[12] A. Augustsson, "A Framework for Evaluating Regression Test Selection Techniques in Industry," 2012.

[13] R. Malhotra and M. Khari, "Test Suite Optimization using Mutated Artificial Bee Colony."

[14] A. Panichella, R. Oliveto, M. di Penta, and A. de Lucia, "Improving multi-objective test case selection by injecting diversity in genetic algorithms," *IEEE Transactions on Software Engineering*, vol. 41, no. 4, pp. 358–383, Apr. 2015, doi: 10.1109/TSE.2014.2364175.

[15] IEEE Singapore Section, IEEE Region 10, and Institute of Electrical and Electronics Engineers, *Proceedings of the 2016 IEEE Region 10 Conference (TENCON) : November 22-25, 2016, Marina Bay Sands, Singapore.*

[16] E. Engström, P. Runeson, and M. Skoglund, "A Systematic Review on Regression Test Selection Techniques."

[17] R. Kazmi, D. N. A. Jawawi, R. Mohamad, and I. Ghani, "Effective regression test case selection: A systematic literature review," *ACM Computing Surveys*, vol. 50, no. 2. Association for Computing Machinery, May 01, 2017. doi: 10.1145/3057269.

[18] A. Choudhary, A. P. Agrawal, and A. Kaur, "An effective approach for regression test case selection using pareto based multi-objective harmony search," in *Proceedings - International Conference on Software Engineering*, May 2018, pp. 13–20. doi: 10.1145/3194718.3194722.

[19] D. K. Yadav and S. Dutta, "Regression test case selection and prioritization for object oriented software," *Microsystem Technologies*, vol. 26, no. 5, pp. 1463–1477, May 2020, doi: 10.1007/s00542-019-04679-7.

[20] N. Gokilavani and B. Bharathi, "Multi-Objective based test case selection and prioritization for distributed cloud environment," *Microprocess Microsyst*, vol. 82, Apr. 2021, doi: 10.1016/j.micpro.2021.103964.

[21] M. V. Mangave, "OPTIMIZATION USING DIFFERENT CRITERION FOR TEST CASE SELECTION IN REGRESSION TESTING."

[22] A. Bajaj and O. P. Sangwan, "A survey on regression testing using nature-inspired approaches," in *2018 4th International Conference on Computing Communication and Automation, ICCCA 2018*, Dec. 2018. doi: 10.1109/CCAA.2018.8777692.

[23] S. Singhal, N. Jatana, B. Suri, S. Misra, and L. Fernandez-Sanz, "Systematic literature review on test case selection and prioritization: A tertiary study," *Applied Sciences (Switzerland)*, vol. 11, no. 24. MDPI, Dec. 01, 2021. doi: 10.3390/app112412121.

[24] R. Nagar, A. Kumar, S. Kumar, and A. S. Baghel, "Implementing test case selection and reduction techniques using meta-heuristics," in *Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit*, Nov. 2014, pp. 837–842. doi: 10.1109/CONFLUENCE.2014.6949377.