Evolutionary Intelligence (2025) 18:62
https://doi.org/10.1007/s12065-025-01046-6

RESEARCH PAPER

Beyond publication numbers: a novel approach to academic ranking using evolutionary programming

Ghulam Mustafa[3] · Muhammad Tanvir Afzal[3] · Abid Rauf[1] · Muhammad Abdullah Khan[2]

Received: 2 November 2024 / Revised: 12 April 2025 / Accepted: 19 April 2025 / Published online: 28 May 2025

Abstract
Evaluating the impact of researchers within a scientific community remains a complex and debated issue. Traditional metrics, such as publication and citation counts, often fail to capture the multifaceted nature of scholarly influence. While several indices, such as the h-index and its variants have been proposed, they too face limitations in reflecting qualitative and longitudinal aspects of academic contributions. This study introduces a novel researcher ranking framework that integrates four distinct parameters: Normalized Citation Score (NCS), Complete Career Contribution (CCC), Temporal Citation Input (TCI), and Collaborative Prestige (CP). These parameters are combined using two ranking mechanisms: the Statistical Ranking System (SRS), based on mathematical models, and the Comprehensive Ranking System (CRS), which employs Genetic Programming (GP) to evolve domain-specific ranking formulas. The methodology was evaluated across five datasets, Computer Science, Mathematics, Neuroscience, Civil Engineering, and a combined dataset, each balanced with equal numbers of awardees and non-awardees. In SRS, models such as the Lehmer Mean and Logarithmic Mean performed effectively in highlighting awardees. In CRS, the evolved models achieved fitness values of up to 0.96 in Computer Science and 0.88 in Mathematics, with slightly lower scores in other domains. The reduced performance on the combined dataset highlights the importance of domain-sensitive modeling. The results suggest that the proposed framework offers a flexible and comprehensive approach to researcher evaluation that can adapt to domain-specific impact patterns, providing an alternative to conventional ranking metrics.

Keywords Author assessment parameters · Parameter ranking · Normalized citation score (NCS) · Complete career contribution (CCC) · Temporal citation input (TCI) · Collaborative prestige (CP) · Arithmetic mean (AM) · Harmonic mean (HM) · Contra-harmonic mean (CHM) · Geometric mean (GM) · Logarithmic mean (LM) · Root mean square (RMS) · Trigonometric mean (TM) · Lehmer mean (LEM)

1 Introduction

At present, the exploration of impactful researchers within a scientific community based on their contributions is a widely debated topic. The assessment of a researcher's scientific impact stands as an essential yet intricate endeavor [1, 2]. Nevertheless, the practice of ranking researchers proves invaluable for the scientific community in making critical decisions, including the selection of candidates for esteemed scientific awards, choosing authors to review papers, disbursing scholarships and grants, assigning tenure positions, and recognizing leading experts in specific fields, among other considerations [3–5]. Furthermore, these rankings offer valuable guidance to students seeking a more pertinent and suitable Ph.D. supervisor [6, 7]. Additionally, governmental bodies can leverage these

Ghulam Mustafa, Muhammad Tanvir Afzal, Abid Rauf, Muhammad Abdullah Khan have contributed to this work.

✉ Ghulam Mustafa
ghulam.mustafa.ssc@stmu.edu.pk

Muhammad Tanvir Afzal
dean.foc@stmu.edu.pk

Abid Rauf
abid.rauf@uettaxila.edu.pk

Muhammad Abdullah Khan
mak30597@gmail.com

1 Department of Computer Science, University of Engineering and Technology, Taxila 47080, Pakistan

2 School of Science, Engineering and Environment SEE Building, University of Salford, Manchester M5 4WT, UK

3 Department of Computing, Shifa Tameer-e-Millat University, Islamabad 44000, Pakistan

rankings to categorize institutions and allocate funds and projects to those whose research profiles occupy top positions [8, 9]. Universities also rely on these rankings for faculty selection, with an organization's website prominently featuring its ranking position, thereby influencing institutional credibility, as highlighted by Mustafa et al. [9, 10], Ahmed et al. [11]. Moreover, students aspiring to pursue studies abroad can leverage these rankings to identify optimal destinations. Similarly, employees can utilize this information to identify top institutions nationally and internationally for employment purposes [12, 13].

To this day, the literature presents an excess of 70 parameters that have been put forth for ranking authors [8, 14]. As outlined by Alshdadi et al. 2023 [15], each parameter employs its distinct criteria in assessing authors, spanning both quantitative and qualitative dimensions, and some exhibiting hybrid nature. Conventionally, authors have been appraised based on factors such as the quantity of publications [16, 17] and citation counts [18]. However, it is widely acknowledged that these metrics alone fail to capture the entirety of a researcher's impact. For instance, multiple researchers may publish in venues with lower quality standards, making publication count an inadequate measure of research impact. Similarly, a high citation count may not reliably indicate the quality, consistency, or longevity of a researcher's work. The prevalence of self-citation and critical citation practices further complicates the assessment of impact.

In response to above mentioned drawbacks, the h-index was introduced by Hirsh, as a composite measure of an author's research output, combining both publication and citation counts [19]. The h-index gained popularity due to its simplicity, efficiency, and ability to encompass both productivity and impact. However, it too has faced criticism within the scientific community. Notably, it does not added the impact of citations received by an author's most highly cited publications and overlooks the collaborative nature of scientific research. To address these concerns, alternative indices such as, the g index [20], t-index [21], AR index [22], e index [23], P index [24], A index [25] etc, have been proposed. Additionally, various studies have been conducted to assess the efficacy of the proposed parameters, aiming to identify the most suitable ones for ranking purposes. These evaluations have been carried out across diverse domains, including Civil Engineering, Mathematics, Computer Science, Neuroscience etc. Despite the plethora of proposed ranking parameters and the extensive evaluative studies undertaken, achieving a unanimous consensus within the scientific community remains challenging [18]. Ongoing research efforts continue in this area, reflecting the persistent quest for a more definitive understanding and agreement [19].

Despite extensive research on evaluating researcher impact, existing bibliometric indices still fall short of providing a comprehensive assessment. Traditional ranking models rely primarily on publication counts, citation counts, and co-author counts, reducing a researcher's entire career to a single numerical value. However, these quantitative metrics fail to capture essential qualitative aspects, such as research consistency, long-term influence, collaboration prestige, and domain-specific variations. Moreover, most ranking methodologies employ static, domain-agnostic formulas, limiting their applicability across different scientific fields. There is no adaptive or learning-based approach that dynamically integrates multiple parameters to optimize rankings based on real-world data. To address these limitations, we propose a novel researcher ranking framework that introduces four distinct parameters:

- Normalized Citation Score (NCS): Captures the quality of research output by considering citation impact across all publications.
- Complete Career Contribution (CCC): Measures the consistency of scholarly output throughout a researcher's career.
- Temporal Citation Input (TCI): Assesses the sustained impact of research over time.
- Collaborative Prestige (CP): Evaluates the influence of a researcher's collaboration network.

To effectively integrate these parameters, we introduce two ranking mechanisms:

- Statistical Ranking System (SRS): Uses advanced statistical models (e.g., Lehmer Mean, Geometric Mean, Logarithmic Mean) to determine the optimal parameter combinations for ranking researchers in different domains.
- Comprehensive Ranking System (CRS): Employs Genetic Programming (GP) to develop adaptive, domain-specific ranking models, optimizing parameter integration based on real-world awardee data.

To validate the proposed methodology, we conducted extensive experiments across five distinct datasets (Computer Science, Mathematics, Neuroscience, Civil Engineering, and a Combined Dataset), ensuring a balanced evaluation with equal numbers of awardees and non-awardees. Our key contributions include:

- Proposing four novel parameters that provide a holistic evaluation of researcher impact, capturing both quantitative and qualitative aspects.

- Compiling domain-specific datasets to ensure fairness and evaluate the effectiveness of the proposed ranking framework.
- Implementing a neural network-based weight optimization method to determine the relative importance of each parameter across different domains.
- Developing a statistical approach (SRS) to combine proposed parameters into a new ranking index using optimal mathematical models.
- Introducing an adaptive ranking model (CRS) based on Genetic Programming (GP) to dynamically evolve domain-specific ranking equations.
- Demonstrating the effectiveness of our ranking framework, where CRS models achieved high accuracy, with fitness values of 0.9 for Computer Science and 0.88 for Mathematics, demonstrating improved performance compared to conventional techniques in our evaluated datasets.

By bridging the gaps in existing bibliometric indices, this study provides a robust, adaptive, and domain-sensitive ranking methodology, ensuring fairer and more accurate researcher evaluation across multiple disciplines.

The remaining documents are organized as follows. The subsequent section offers an overview of previously proposed techniques, followed by the presentation of the proposed methodology in the subsequent literature review. The results section outlines the experimental findings. The conclusion section summarizes the proposed study, while the final section delves into potential avenues for future research.

2 Literature

Assessing the scientific output of authors holds considerable significance within diverse research-oriented areas. This task serves a crucial role in informed decision-making, identifying best persons for scientific accolades, aligning authors with fitting research endeavors, helping in career advancements, bestowing well-deserved tenures, and contracting proficient experts in the field [26]. Over the past two decades, the literature has proposed more than 70 parameters to evaluate and ranked authors based on their research contributions [27]. Among the earliest metrics used for evaluating researchers is the total no of publication count [28]. However, the publication count parameter has limitations in capturing both the impact and quality of an author's work. Some researchers attempt to increase their total number of publications by submitting articles to low-impact journals, which compromises the reliability of this measure. In response to this constraint, the introduction of citation count became pivotal. Nonetheless, citation count has its own set of drawbacks. Newly published papers often require time to accumulate citations, which can disadvantage new authors, even if they have a substantial number of publications. Moreover, the total citation count may not always provide an accurate measure of publication quality, as researchers may cite papers primarily for critique rather than endorsement. Additionally, some authors engage in self-citation, referencing their own work to artificially inflate their citation count.

To address the drawbacks of relying solely on the total number of publications and citations, Hirsch [19] proposed the h-index as an alternative metric. This quickly gained prominence and became a key parameter for evaluating a researcher's productivity. Nevertheless, this parameters also not exempt from limitations. Its reliance on long-term observations results in older authors typically having higest h-index values compared to their newer competitors, and its applicability is effected by the specific domain research. Furthermore, the h-index is less sensitive to the increase in citations for a researchers core articles, which may not significantly influence the overall h-index. In response to these challenges, multiple alternative indices have been introduced, such as the g index [20], t-index [21], AR index [22], e index [23], P index [24], A index [25], m-quotient [29], contemporary h-index [30], f-index [31], Wu-index [32], and q2-index [28], among others. Notably, the main drawback of these novel techniques is that, upon their proposition in the literature, they are typically tested in hypothetical or fictional case scenarios [33].

Multiple research studies have systematically explored various author assessment parameters across multiple domain datasets, contributing valuable insights into their effectiveness. For instance, Kosmulski et al. [27] delved into the relationship between two indices, h-index and h(2) index, analyzing data of 19 professors which are faculty members of University of Poland in chemistry department. Their findings presented a robust correlation in between these indices, which portraying closely aligned outcomes and similar results trend. Similarly, Van et al. [34] conducted a comprehensive study comparing the h-index and some of its variants, examining data which are collected from 147 research groups lie in the chemistry department in the Netherlands. Unlike many research studies focusing on individual performance, this research specifically evaluated the performance of research groups, considering a three-year citation window instead of an researchers lifetime citations. Jin et al. [22] proposed a new approach in their study, integrating multiple indices to evaluate authors. Their method combined two indices one quantitative and the other qualitative including metrics such as the h-index, r-index, and ar-index. The results highlighted that the combination of the h-index and ar-index effectively indicated research performance. In 2008, Schreiber [35] conducted a study evaluating the g-index alongside other indices such as the h-index, r-index, and a-index. The analysis was based on data from

26 physicists from the Physics Department at Chemnitz University. Schreiber's findings suggested that the g-index was a more reliable parameter for assessing the overall impact of a researcher's publications compared to the h-index. In 2016, Xiao et al. [36] conducted a comprehensive evaluation of 29 h-index variants, analyzing their correlations with benchmark indices such as the h-index and Wu index. The study aimed to determine the degree of correlation each variant had with these benchmarks. The results indicated that indices highly correlated with the h-index tended to show lower correlations with the Wu index, and vice versa. Furthermore, the study found that highly correlated indices offered only marginal improvements over the h-index or Wu index.

In 2016, Ayaz et al. [37] conducted a comprehensive study evaluating the complete-h index using a dataset of awardees from a well-known mathematical scientific society. The results demonstrated the superior performance of the complete-h index in this specific context, outperforming other indices. Similarly, in 2018, Raheel et al. [6] assessed the h-index and its variants, incorporating publication age and citation intensity within the field of civil engineering. The findings indicated that the Wu index emerged as the most effective measure compared to other indices. Moving into 2019, Ghani et al. [38] examined the h-index and some of its variants within the domain of mathematics. Notably, their study highlighted the fraction count on paper parameter as a top performer, accurately elevating 55 percent researchers who won awards within the top 10 records. In the same year, Ameer and Afzal [39] scrutinized quantitative parameters employing a dataset which contains researchers belonging from neuroscience domains. Their investigation underscored the effectiveness of the R-index and hg-index in elevating awardees at the pinnacle positions. Concurrently, Ain et al. [40], in another 2019 study, sought to evaluate author assessment parameters by employing the dataset which contain researchers belonging to mathematics domain. They aimed to establish the correlation between the h-index and some of its variants for researchers, ranking these parameters based on their association with award-winning researchers. Fast forwarding to 2021, Salman et al. [30] concentrated on multi-authorship indices, using a dataset which contain researchers belonging to civil engineering domains. Their study indicated that the gf index performed extraordinary as compare to all other parameters discussed in this study. Nonetheless, a common limitation in these studies was their effort to link the h-index or its variations to researchers who had already received awards prior to the introduction of these metrics. This suggests that the award recipients might not have depended on the h-index or its variants, making any observed correlations potentially coincidental. To address this limitation, Usman et al. [41] conducted a study evaluating the h-index and some of its variants using data from civil engineering researchers. The selection process included both randomly chosen and purposefully selected researchers, encompassing both awardees and non-awardees from the same period, with a particular emphasis on those who received awards after 2005. This comprehensive approach provided more definitive insights into the key parameters for identifying award winners. In a recent development, Abdulrahman et al. [15] introduced a rule for the scientific community using deep learning models. They analyzed datasets from multiple domains, including civil engineering, mathematics, and neuroscience, incorporating data from 500 researchers in each field. Their findings demonstrated that these rules achieved a high accuracy rate of up to 70%. Mustafa et al. conducted two studies [16, 18]. The first study evaluated publication and citation count-based parameters, highlighting that the normalized h-index outperformed all other parameters. The second study focused on publication age-based parameters and found the Ar index to be the best among the indices considered, using a mathematics domain dataset for experimental purposes. Furthermore, Bilal et al. [1] evaluated author count-based parameters using a similar dataset, reporting that the hf index outperformed all other indices in this context. Furthermore, in 2024, Mustafa et al. conducted four additional studies in this domain [42–45]. Initially, they proposed the MERT Technique for ranking author assessment parameters, using datasets from mathematics and civil engineering in two separate studies [42, 43]. These studies reported that the normalized h-index achieved the highest importance score in both domains. Additionally, they identified the Trigonometric Mean as the best statistical model for combining author assessment parameters. In another study [44], they introduced a new index called the GK index, which combines the Gf and K indices using the geometric mean. This novel index improved results by 12% compared to existing indices. In their final study [45], they proposed a set of rules using association rule mining for the mathematics domain. Following these rules could help researchers become top authors in the future. Table 1 presents a summary of some recent studies related to this work.

2.1 Critical analysis

After conducting a thorough review of the available literature and based on the recent studies discussed in Table 1, we have observed that the studies in the literature fall into two categories: the proposition of indices and the evaluation of existing indices. From both categories, we have identified the following gaps:

- The proposed indices are primarily evaluated in hypothetical scenarios.
- These indices focus mainly on publication count, citation count, or co-author count. None of them have the

Table 1 Summary of recent studies of related work

Reference (Year)	Key Idea	Strengths	Weaknesses
[39] (2019)	Evaluation of h-index and its qualitative and quantitative variants.	Used a benchmark dataset to evaluate author assessment parameters.	Considered only a few parameters for evaluation and focused solely on the neuroscience domain.
[40] (2019)	Evaluation of author assessment parameters.	Used a benchmark dataset to evaluate author assessment parameters.	Limited the evaluation to a small number of parameters, validated only in the mathematics domain.
[31] (2021)	Assessment of author ranking indices based on multi-authorship.	First evaluation to consider multi-authorship parameters.	Focused on only a few parameters, validated in a single domain.
[41] (2021)	Evaluation of existing author assessment parameters.	Employed Logistic Regression techniques for ranking, a novel approach in this context.	Utilized only a limited number of author assessment parameters.
[8] (2023)	Review of h-index a nd its alternative indices.	Discussed over 70 parameters, detailing their pros and cons.	Parameters reviewed were primarily based on total publications, total citations, or the number of co-authors.
[18] (2023)	Comprehensive evaluation of publication and citation metrics for quantifying scholarly influence.	Included all parameters related to publication and citation counts.	Focused exclusively on the mathematics domain.
[15] (2023)	Guidelines for researchers using rule-based approaches.	Utilized a well-known mechanism, such as Decision Trees, for generating rules.	Analysis was based on only a limited number of author assessment parameters.
[42] (2024)	Proposed the MERT technique for ranking author assessment parameters.	Applied a machine learning algorithm for calculating impact scores, avoiding manual approaches.	Ranking was performed on a single domain dataset.
[43] (2024)	Proposed comprehensive rules to guide researchers toward becoming influential.	Rules were built on top-ranked parameters, providing actionable guidance for researchers.	Rules are specific to the mathematics domain and may not generalize to other fields.
[44] (2024)	Proposed a new index, "GK index."	*Conducted a thorough analysis before introducing the index.*	The index is valid for only one domain and may fail with different datasets. It heavily relies on publication count and the number of authors per publication.
[45] (2024)	Proposed a ranking mechanism for author assessment parameters.	*Used a large number of author assessment parameters for ranking.*	Limited the analysis to a single domain dataset.

ability to capture or measure the lifetime achievements of a researcher.

- Evaluation studies have assessed these proposed indices, but no study has reported a generalized index applicable to a specific domain or across domains, as each study has evaluated indices on a particular domain dataset.
- Some studies have used statistical models to combine existing indices, but these are limited, and a more comprehensive exploration is required.
- No alternative approaches have been proposed to combine indices and develop a new one.

To address these gaps, we gathered a comprehensive dataset from multiple domains, including computer science, mathematics, neuroscience, and civil engineering. Additionally, we propose new parameters that address a significant gap in the assessment of a researcher's entire career. It has become evident that the parameters commonly outlined in the literature often fail to provide a holistic understanding of a researcher's lifelong contributions. These parameters typically include publication count, citation count, and co-author count. However, the prevailing practice is to distill these multifaceted aspects into a single numerical value, whether it be publication count, citation count, co-author count, or a combination of these. Unfortunately, such approaches often overlook the broader perspective necessary to fully understand a researcher's lifetime achievements. Therefore, we propose a paradigm shift in the evaluation process. Instead of assessing researchers based on individual or combined parameters, we advocate for the consideration of a unified metric. This comprehensive metric should not only account for publication count, citation count, and co-author network, but also the prestige associated with the network they have cultivated throughout their career. In essence, our recommendation is to adopt a more encompassing and nuanced approach that captures the breadth and depth of a researcher's contributions over the entirety of their professional journey. Furthermore, after proposing these new parameters, we introduce two methods for combining them to develop a new index that integrates these factors and provides a more complete assessment of a researcher's lifetime achievements.

2.2 Metaheuristic optimization algorithms

Metaheuristic optimization algorithms have gained significant attention in solving complex optimization problems, particularly in machine learning, artificial intelligence, and evolutionary computation. These algorithms offer adaptive mechanisms to explore and exploit the search space efficiently, making them well-suited for problems involving multi-objective optimization, large-scale datasets, and computational constraints. In the context of researcher ranking, optimization plays a crucial role in parameter weight selection, model tuning, and ranking system improvement.

Several recent metaheuristic algorithms have been introduced to enhance optimization performance across diverse domains. The Potter Optimization Algorithm (POA) [46] is inspired by the problem-solving behavior of Harry Potter's spell casting and decision-making strategies. Similarly, the **Carpet Weaving Optimization (CWO)** [47] draws inspiration from the intricate process of Persian carpet weaving, emphasizing pattern formation and iterative refinement.

Bio-inspired algorithms have also emerged as effective optimization strategies. The Fossa Optimization Algorithm (FOA) [48] mimics the hunting behavior of fossas, enhancing solution exploration and convergence speed. Likewise, the Addax Optimization Algorithm (AOA) [49] is based on the adaptive survival strategies of addax antelopes in harsh desert environments, making it robust for high-dimensional optimization problems.

Several recent metaheuristics are inspired by human activities. The Dollmaker Optimization Algorithm (DOA) [50] simulates the fine-tuning process of crafting dolls, balancing precision and randomness. Similarly, the culptor Optimization Algorithm (SOA) [51] is inspired by sculpting techniques, incorporating iterative refinement and structural adaptation. The Sales Training Based Optimization (STBO) [52] models training-based sales strategies, where solution refinement mimics the improvement of sales skills over time.

Nature-inspired techniques have also demonstrated effectiveness. The Orangutan Optimization Algorithm (OOA) [53] is modeled on the intelligent problem-solving behavior of orangutans, integrating adaptive learning into optimization. The Tailor Optimization Algorithm (TOA) [54] simulates the precision and iterative adjustments made in tailoring, enhancing fine-tuned search mechanisms. Additionally, the Spider-Tailed Horned Viper Optimization (STHVO) [55] is inspired by the predatory techniques of the spider-tailed horned viper, balancing exploitation and exploration for optimal performance.

These recent advancements in metaheuristic optimization demonstrate their potential in refining researcher ranking models. Given that our study incorporates Genetic Programming (GP) for ranking optimization, future extensions may explore the integration of these cutting-edge algorithms to further improve ranking accuracy and parameter tuning.

3 Methodology

After conducting an exhaustive literature review, our focus shifts towards introducing a Comprehensive Ranking System (CRS) for assessing researchers. Figure 1 illustrates the architectural framework, outlining distinct stages comprising: (i) Domain Selection, (ii) Dataset Collection, (iii) Calculate Proposed Parameters, (iv), Statistical Analysis and (v) Genetic Programming (GP). The subsequent sections provide an in-depth exploration and explanation of each of these steps.

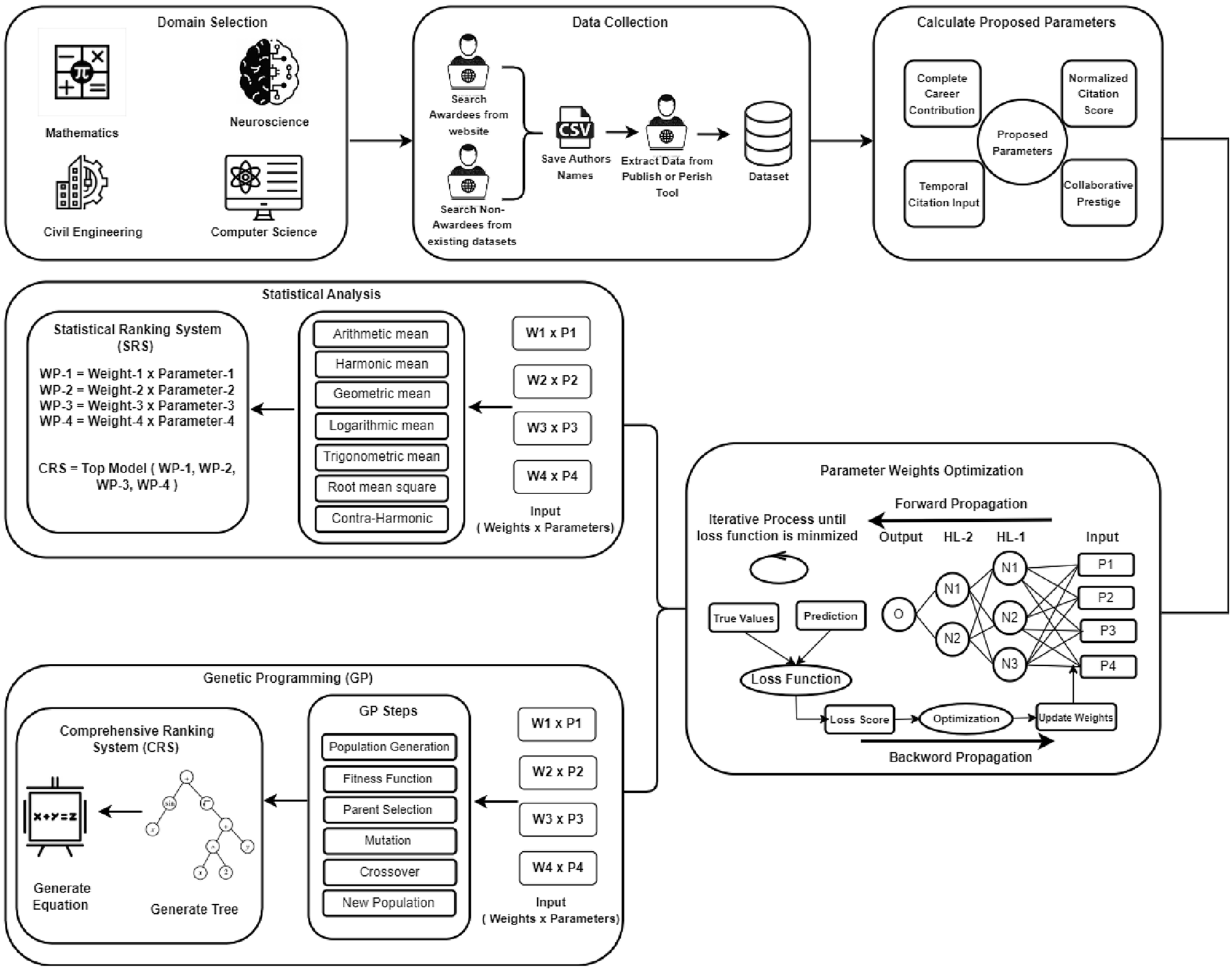

Fig. 1 Architecture diagram of proposed methodology

3.1 Domain selection

In this research study, the careful selection of specific domains-Computer Science, Neuroscience, Mathematics, and Civil Engineering-holds paramount significance. Each field was chosen based on distinct reasons, ensuring a comprehensive and diverse exploration of indices and methodologies in our research. The inclusion of the Computer Science domain is justified by its rapidly evolving nature, marked by continuous innovation and advancements. Engaging in research in this domain allows contributors to play a pivotal role in developing new technologies, methodologies, and solutions. This rapid progress and improvement span across various sectors, underscoring the dynamic nature of the field. Civil Engineering, being one of the oldest subjects in human history, has witnessed extensive research. This rich history positions it as an ideal field for evaluating the proposed methodology. Despite its fundamental importance, authors in this domain have not been thoroughly analyzed to define a standard awarding criterion. The proposed methodology holds promise for identifying deserving individuals in Civil Engineering, thereby supporting their continued development and growth. Additionally, given the significance of Civil Engineering in the modern construction era, further research in this area is imperative. Mathematics, considered a core pillar for various branches of sciences, was selected due to its intrinsic connection to fields like physics, computer science, and chemistry. The interdisciplinary nature of mathematics, with connections to multiple scientific disciplines, positions it as a vital domain for the study, demonstrating its relevance across diverse fields. The selection of the Neuroscience field is motivated by the substantial research being conducted in this domain. Neuroscience explores diseases, disorders, and injuries related to the nervous system, with researchers striving to develop methods for preventing and treating issues that impact the brain, nervous system, and body. The extensive research in this field highlights the importance of ranking researchers as a significant endeavor.

3.2 Datasets

Based on the selected domains (discussed in the previous section), a dataset comprising data of researchers, including awardees and non-awardees, has been curated. For the field of Civil Engineering, awardees were identified from prestigious scientific societies such as ACI, ASCE, CSCE, and ICE, spanning the last three decades. Non-awardees were sourced from the dataset of Usman et al. [41]. In the realm of neuroscience, a list of non-award-winning researchers was compiled from Ameer and Afzal. [39] and extracted the names of award winners from reputable scientific societies like SFN, FENS, CNS, and ANS. The dataset of Mustafa et al. [10] was utilized to create a list of awardees and non-awardees for mathematics. Meanwhile, for mathematics award recipients, exploration of IMU, AMS, and LMS websites was undertaken. Computer Science data, utilizing the Artminor dataset from Ayaz et al. [37, 66] for non-awardees and information from prestigious scientific awards such as IEEE Fellow and Turing for award winners, was also included. A detailed discription of dataset's statistics is provided in Table 2. The data collection process involved visiting various scientific society websites within the fields of mathematics, neuroscience, civil engineering, and computer science over the past three decades. The Publish or Perish platform was used for data extraction, employing a 'hold on' strategy to collect authors' data from both before and after their award-winning years. To maintain balance, an equal number of awardees and non-awardees were collected for each year. Following this, a meticulous data cleansing process was conducted to ensure accuracy and remove duplicate entries. Two key processes were implemented to enhance data quality: a consistency filter for research articles and an author disambiguation process to detect and eliminate duplicate entries from researchers publishing under different name variations.

Table 2 Dataset Statistics

Mathematics	
Authors Count	1050
Awardees Count	525
Non-awardees Count	525
Publications Count	204,896
Citations Count	14,370,007
Civil Engineering	
Authors Count	1180
Awardees Count	590
Non-awardees Count	590
Publications Count	214,672
Citations Count	24,061,210
Neuroscience	
Authors Count	1060
Awardees Count	530
Non-awardees Count	530
Publications Count	166,871
Citations Count	25,855,493
Computer Science	
Authors Count	1200
Awardees Count	600
Non-awardees Count	600
Publications Count	171,388
Citations Count	32,801,476

3.3 Proposed parameters

In the realm of academic evaluation, conventional metrics used to assess researchers' contributions often fail to fully capture the depth and breadth of their lifetime achievements. The common practice of reducing a researcher's impact to singular numerical values, such as publication counts or citation metrics, oversimplifies the multifaceted nature of academic pursuits. In response to this limitation, our research advocates for a paradigm shift in how we evaluate researchers' lifelong impact. Rather than relying solely on numerical indices like publication counts, citation counts, co-authorship metrics, h-index, g-index, and others, this study propose a focus on qualitative attributes that span an entire research career. These qualitative attributes encompass the quality, consistency, reputation, and prestige of a researcher within the scientific community. Our contribution lies in identifying and quantifying these parameters using logical and philosophical reasoning, which, to our knowledge, has not been adequately addressed in existing literature. Therefore, our research objective is to provide a means of quantifying these qualitative measures, allowing for a more comprehensive evaluation of researchers' contributions. To achieve this objective, a set of four distinct parameters is proposed, collectively offering a more holistic and nuanced evaluation of the aforementioned qualities. These parameters move beyond mere quantification, aiming to capture the quality, consistency, popularity, and collaborative prestige inherent in academic endeavors. This section outlines these proposed parameters-Normalized Citation Score, Complete Career Contribution, Temporal Citation Input, and Collaborative Prestige-each crafted to provide a comprehensive perspective on a researcher's enduring impact and contribution throughout their academic journey.

3.3.1 Normalized citation score (NCS)

$$NCS = \frac{\sum_{i=1}^{n} T_{C_i}}{T_p} \tag{1}$$

Where NCS is the Normalized Citation Score, T_{C_i} is the Citation of the individual publication of the author and T_P is the Total publication of the authors.

The NCS parameter has been introduced to quantify the quality of the researcher work. This parameter combines publications and citations to overcome limitations in existing author indexes. Publication quality is typically measured by citations, but current indexes only focus on top-cited papers, excluding highly cited ones. For example, the h-index selects the top h-papers with at least h-citations, ignoring citations and publications below h, as well as papers with more than h-citations. Similarly, the g-index considers highly cited papers but disregards tail citations and papers. Several other parameters with similar limitations exist in literature. To our knowledge, no one has considered a researcher's total citations across all publications. NCS addresses this gap by considering all citations and publications, maintaining the quality of research throughout a researcher's career. The strength of NCS lies in its ability to credit early researchers for quality work without waiting for extensive experience. For instance, one paper with 50 citations may give an h-index of 1 but an NCS value of 50. This value, however, decreases if the researcher stops producing quality work. Established researchers benefit from NCS by receiving preferences for quality work and facing penalties for a decline in research quality. NCS offers advantages for both early-career and experienced researchers, unlike other parameters. Early-career researchers benefit by being recognized as influential from their first paper if its citation count equals its N value. Established researchers benefit as NCS compensates for variations in citation patterns, providing a more balanced assessment of their impact over time. This parameter assesses the researcher's quality in scenarios where papers receive varying citation counts. NCS elevates high-quality researchers and adjusts their parameter value based on their ongoing contributions. If a person publishes a highly cited first paper followed by a less-cited second paper, their NCS value decreases, emphasizing the importance of consistently producing impactful and quality research to maintain a high parameter value.

3.3.2 Temporal citation input (TCI)

$$TCI = \frac{\sum_{i=1}^{n} T_{C_i}}{T_y} \tag{2}$$

Where TCI is the Temporal Citation Input, T_{C_i} is the Citation of the individual publication of the author and T_y is the Total Years of author works.

The next parameter, TCI, has been introduced to quantify the consistency of reputation of the researchers. This parameter integrates the total years of an author's work with the total number of citations received, with the objective of assessing the consistency of a researcher's popularity within their peer or contemporary scientific community. The TCI parameter, however, relies solely on the total number of publications and their corresponding citation counts, disregarding the fact that scholars may exhibit varying levels of productivity at different stages of their careers. For instance, consider a retired or less active researcher whose previously published articles continue to garner regular citations; such an individual rightfully maintains prominence despite a reduced output of new publications. Conversely, a young scholar with a relatively shorter career may have published only a few papers, yet their lower citation count does not accurately reflect their potential impact. Recognizing this disparity, the TCI parameter emerges as a viable solution. Existing indices attempting to address this issue, such as the AR index, primarily focus on core article citations, while others like the timed h-index, hpd index, and ACPD index concentrate on specific time frames, such as the last 2 or 3 years. However, these time-bound metrics fail to encapsulate the overall impact of a scholar's body of work. Notably, no prior index has systematically considered a researcher's total citations across all the years of their career. Hence, the TCI parameter proposed in our research aims to ascertain whether the number of citations increases consistently over time, remains stable, or decreases. To attain a X TCI value, an individual must consistently accumulate X citations each year until the end of their career. This implies that the researcher consistently engages the academic community or attracts citations through continued high-quality work. Conversely, a decline in reputation over the years would result in a decrease in the TCI value, possibly reaching near-zero at a certain point.

3.3.3 Complete career contribution (CCC)

$$CCC = \frac{T_p}{T_y} \tag{3}$$

Where CCC is the Complete Career Contribution, T_p is the Total Publication of the author and T_Y is the Total Years of author works.

The CCC parameter has been introduced to quantify the consistency of researchers in their scholarly output. Unlike previously proposed parameters, CCC specifically focuses on the cumulative output of scholarly work since the inception of a researcher's career journey. As per our knowledge in literature no one addressed the total publication over the total years. The primary objective of the CCC parameter is to rank researchers based on their sustained commitment and consistency in producing scholarly work. It aims to distinguish researchers who consistently contribute to their field over time from those who achieve success through sporadic bursts of productivity. For instance, consider a researcher who attained an h-index of 150 ten years ago and maintains this score without publishing any new articles in the subsequent decade. While their h-index remains static, it does not accurately reflect their ongoing commitment to research. Thus, the CCC parameter seeks to assess the enduring dedication of researchers to their work. In the calculation of the CCC parameter, as the number of years' increases, so does the expectation for the researcher's publication output. To achieve a X CCC value, an individual must consistently publish a minimum of X papers every year until the end of their career. If the researcher's commitment wanes, leading to a decrease in publications over time, their CCC value will correspondingly decline. Conversely, actively increasing scholarly output in preceding years can elevate and improve the CCC parameter. In essence, the CCC parameter serves as a dynamic measure of a researcher's consistency and commitment to their field, reflecting their ongoing contributions over the course of their career.

3.3.4 Collaborative prestige (CP)

$$CP = \frac{C_H}{T_A} \tag{4}$$

Where CP is the Collaborative Prestige, C_H is the Co-author h-indices and T_A is the Total number of Co-authors.

The CP parameter has been introduced to quantify the prestige of a researcher, drawing inspiration from the well-known adage, "A man is known by the company he keeps." This parameter pertains to the researchers with whom an individual collaborates or associates, a dimension that has not been addressed in the existing literature. In essence, the CP parameter seeks to evaluate a researcher's standing within the academic community based on their professional relationships and collaborations. For instance, if a researcher consistently collaborates with individuals who possess high h-index scores or are esteemed figures in their field, it reflects positively on their own prestige and standing within the community. This could indicate that they have earned the respect and recognition of established scholars, either through their own contributions or by being invited to participate in collaborative projects. This parameter serves to assess the dynamic engagement of a researcher within the scientific community in terms of co-authorships and collaborative endeavors. It recognizes that the company a researcher keeps can significantly impact their reputation and influence within their field. Researchers with a high CP value are likely to have positioned themselves within a network of esteemed colleagues and pioneers in their respective disciplines. Overall, the CP parameter provides valuable insights into the professional networks and collaborations of researchers, shedding light on their prestige and influence within the academic community. By considering the quality of their collaborators and the nature of their professional relationships, this parameter offers a nuanced perspective on a researcher's standing and reputation within their field.

3.4 Parameter weight optimization

Before formulating a new comprehensive index by combining the proposed parameters, it is imperative to ascertain the relative importance of these parameters in identifying influential authors. In many machine learning problems, not all features hold equal weight in making predictions or classifications [56]. Certain features may exert a more pronounced influence on the output compared to others. Consequently, the scientific community has increasingly embraced the assignment of weights to these parameters, reflective of their significance. The methodology for determining these weights is explained by Neural Networks [57]. Neural networks utilize weights to assign varying degrees of importance to distinct features. Throughout the training phase, the network iteratively adjusts these weights to minimize the disparity between predicted and actual outputs. This adjustment process is facilitated by the optimization algorithm, which modifies the feature weights based on the error computed during training. Features that contribute more substantially to error reduction are assigned higher weights.

To accomplish this, the Feature Weighting technique introduced by Zeng and Martinez was employed [58]. This approach centers on extracting information regarding the significance of features from a trained neural network. Initially, the model trains a multilayer neural network utilizing the backpropagation algorithm. During training, the learning rate was set to 0.2, and the momentum was set to 0.5. The dataset D is randomly divided into two subsets: a training set $D1$ (two-thirds of D) and a validation set $D2$ (one-third of D). The neural network is trained using $D1$, while $D2$ is used to evaluate its performance to prevent overfitting. The network is designed with a single hidden layer. The number of hidden nodes H is dynamically determined by monitoring the accuracy on $D2$. Initially, H is set to 1. After training with a fixed H, the accuracy on $D2$ is recorded. H is incremented by 1 until the optimal H is found based on the criterion that the accuracy does not improve with further increases in H. Once the optimal H is identified, the entire dataset D is used to train the network with this fixed H, starting from the saved weights for the best H. For each fixed H, accuracy on $D2$ is monitored every 10 epochs. If no improvement in accuracy is observed after 200 epochs, training for that H is stopped. Finally, feature weights are extracted from the trained network by determining the weight for each input node i.

$$W_i = \sum_{J=1}^{H} \sum_{K=1}^{O} |V_{i,j} * V_{j,k}| \tag{5}$$

In the given equation, W_i denotes the feature weight for the input node i; $V_{i,j}$ represents the network weight (link strength) from the input node i to the hidden node j, and $V_{j,k}$ indicates the weight from the hidden node j to the output node k. Here, H stands for the number of hidden nodes and O signifies the number of output nodes. Each term in Eq. (1) represents a path from input node i to output node k through hidden node j. The summation accounts for all possible forward paths from each input node to every output node. The purpose of Eq. (5) in weighting features is to highlight significant features by measuring their overall influence on the output nodes through the hidden nodes. This influence is reflected in the strengths of the links across these paths. Equation (5) is used to quantify this influence. After computing the weights for all features using Eq. (5), the weights are normalized to ensure their sum equals 1. This normalization is achieved using Eq. (6):

$$W_i = \frac{W_i}{\sum_{i=1}^{N} W_i} \tag{6}$$

where W_i represents the normalized weight for feature i, ensuring that the total sum of the normalized weights is equal to 1.

After determining the parameter weights, two indices are proposed for combining these parameters. The first index is derived through a Statistical Algorithm, while the second index is based on Genetic Programming.

3.5 Statistical ranking system (SRS)

Statistical techniques play a pivotal role in data analysis across various research fields (Tzenios, 2023).To develop the Statistical Ranking System (SRS), advanced statistical models were utilized to integrate the four proposed parameters. These techniques include the arithmetic mean, contra-harmonic mean, geometric mean, harmonic mean, Lehmer mean, logarithmic mean, root mean square (RMS), and trigonometric mean. These methods are employed to gain a comprehensive understanding of the proposed parameters and evaluate their significance within the dataset using statistical models. The detailed calculations for these methods are provided in Table 3.

In this study, the specified statistical methods were utilized (as detailed in Table 3)) to explore the best approach for integrating the four proposed parameters. By applying these methods, the corresponding values for each researcher were calculated, resulting in unique lists for each statistical approach. Our thorough analysis revealed which statistical model produced the highest number of awardees among the top 100 records. Additionally, these lists were compared to identify the most impactful statistical method for effectively combining the proposed parameter values. The pseudocode for SRS is presented in Algorithm 1

Table 3 Statistical Methods

Method Names	Formulas
Arithmetic mean	$Arithmetic = \frac{X_1+X_2+....+X_n}{n}$
Harmonic mean	$HarmonicMean = \frac{n}{\sum_{i=1}^{n} \frac{1}{X_i}}$
Contra-harmonic mean	$Contra - HarmonicMean = \frac{(X_1^2)+(X_2^2)+...+(X_n^2)}{(X_1+X_2+X_3...+X_n)}$
Geometric mean	$GeometricMean = (X_1 * X_2 * X_3 * ... * X_n)^{\frac{1}{n}}$
Logarithmic mean	$LogarithmicMean = (\frac{1}{n}) * (\sum_{i=1}^{n} log(X_i)))$
Root mean square	$RootMeanSquare = \sqrt{\frac{x_1^2+x_2^2+...+x_n^2}{n}}$ square=
Trigonometric mean	$TrigonometricMean = \frac{\prod_{i=1}^{n} sin(x_i)}{\prod_{i=1}^{n} x_i}$
Lehmer Mean (LEM)	$L_p(x) = \frac{\sum_{k=1}^{n} x_k^p}{\sum_{k=1}^{n} x_k^{p-1}}$

Algorithm 1 Pseudocode for Statistical Ranking System (SRS)

Require: Dataset with four parameters (P_1, P_2, P_3, P_4) for n researchers
Ensure: Ranked list of researchers based on statistical integration of parameters

Step 1: Initialize Rankings
Create an empty dictionary `Rankings` to store researcher scores
Step 2: Compute Scores for Each Researcher
for each researcher i in the dataset **do**
  Compute scores using the following statistical methods:
    • Arithmetic Mean
    • Harmonic Mean
    • Contra-Harmonic Mean
    • Geometric Mean
    • Logarithmic Mean
    • Root Mean Square
    • Trigonometric Mean
    • Lehmer Mean (with suitable p)
  Store the computed scores for researcher i
end for
Step 3: Generate Ranked Lists
for each statistical method **do**
  Generate a ranked list of researchers based on their computed scores
  Save the ranked list
end for
Step 4: Compare Ranked Lists
Identify the statistical method that produces the highest number of awardees in the top 100
Return:
  - The most effective statistical method
  - The final ranked list

The Statistical Ranking System (SRS) integrates four proposed parameters to generate a ranked list of researchers using a systematic approach. Initially, scores for each researcher are calculated using multiple statistical methods, including Arithmetic Mean, Harmonic Mean, Contra-Harmonic Mean, Geometric Mean, Logarithmic Mean, Root Mean Square, Trigonometric Mean, and Lehmer Mean with a suitable parameter value. These computations produce unique scores for each researcher for every statistical method. Based on these scores, ranked lists are generated for each method, which are then evaluated and compared to determine their effectiveness. The method that produces the highest number of awardees in the top 100 records is identified as the most impactful. This systematic evaluation ensures a comprehensive integration of parameters, aiding in the identification of an optimal ranking approach.

3.6 Comprehensive ranking system (CRS) using genetic programming (GP)

Genetic Programming (GP) is an evolutionary algorithm used to evolve symbolic expressions to solve specific problems [59, 60]. In this study, GP is employed to develop mathematical models that represent relationships between academic metrics in the ranking system. GP's tree-like structure is particularly useful for this purpose, as it enables the exploration of complex relationships between input features and the desired outputs. The GP process starts with a randomly generated population using terminal and function sets, which define the search space. A fitness function evaluates the performance of individuals, and the best candidates are selected for reproduction. This process involves crossover and mutation, which help find optimal solutions and introduce diversity in the population. In the context of our work, GP is used for symbolic regression to discover formulas that best represent the relationships between the

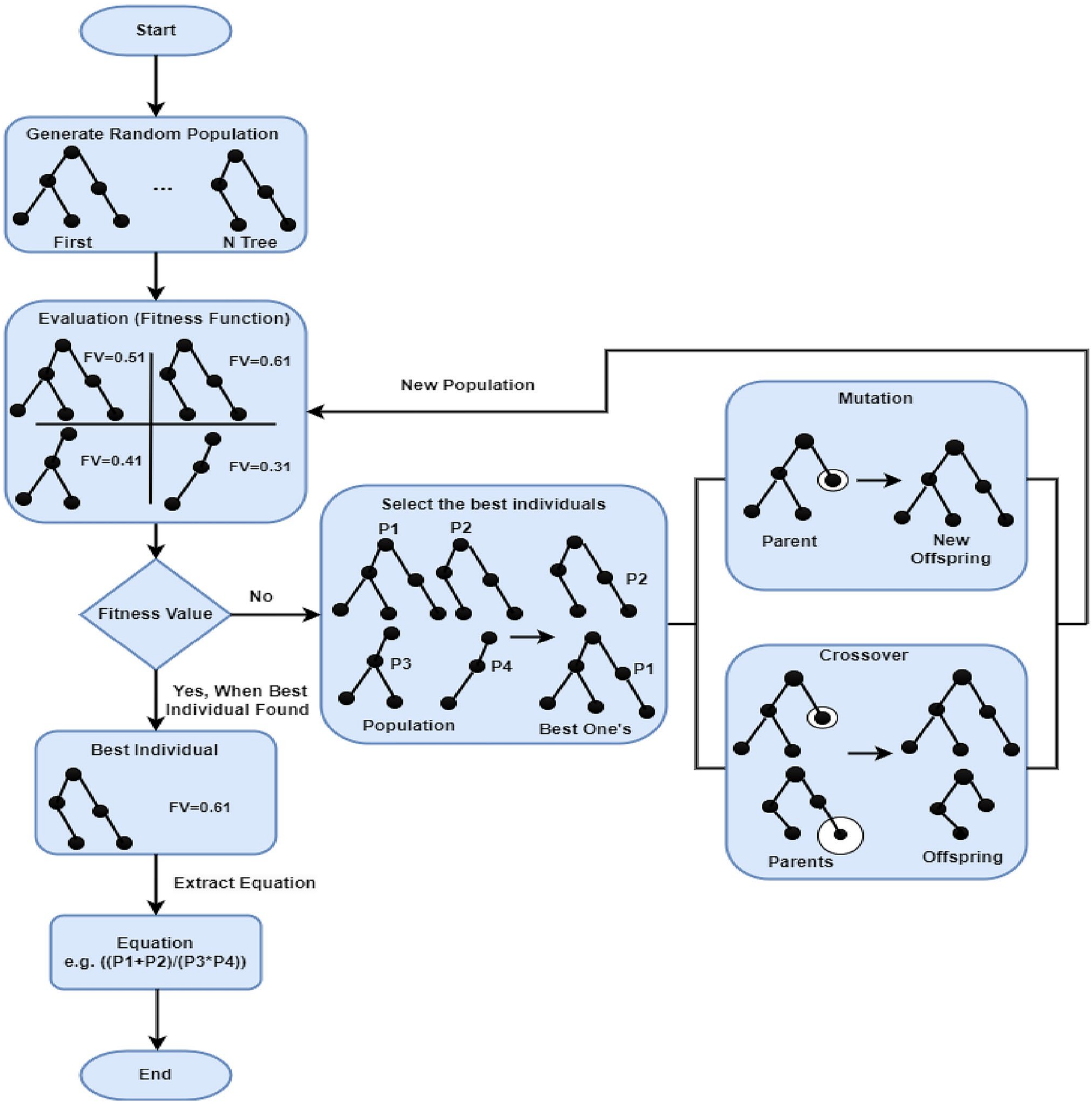

Fig. 2 Flow diagram of GP

ranking metrics and the top 100 records. The evolutionary process helps identify the most accurate expressions for ranking individuals based on their academic performance. The Fig. 2 represents the Genetic Programming technique: The following sections explain the complete methodology of the Genetic Programming mechanism implemented in this study.

3.6.1 Generate random population

After selecting an appropriate representation, the initial population (generation zero) is generated using the ramped half-and-half method. This method creates diversity in the population by generating half of the individuals with the grow method and the other half with the full method. In this study, a population size of 50 was chosen based on the problem's solution space. The function set includes arithmetic operations (*, /, +, -), and the terminals include Normalized

Citation Score (NCS), Temporal Citation Input (TCI), Complete Career Contribution (CCC), and Collaborative Prestige (CP). This approach ensures a varied population that enhances the search for optimal symbolic expressions.

3.6.2 Fitness calculation

This section discusses the concept of fitness, which is essential for guiding the GP algorithm towards optimal solutions. The fitness function used in this study evaluates individuals based on their ability to map input variables to desired output values. Specifically, for each individual, the fitness value is determined by counting how many positive cases are returned within the top 100 records, reflecting the individual's effectiveness in prioritizing relevant outcomes. The fitness cases consist of input-output pairs, and the fitness function steers the algorithm toward solutions that best match these cases. In this study, the fitness function is designed to evaluate the performance of individuals in identifying the most relevant results, crucial for the task at hand.

$$FitnessCases(FC) = Sort(Input, Output) \tag{7}$$

$$FitnessValue = \sum_{i=1}^{100} Count(F_{C_i} - Output == 1) \tag{8}$$

Where i represent the individual fitness case and FC_i-Output represent the individual fitness case output.

3.6.3 Selection method

Selection methods in genetic programming (GP) are pivotal for choosing individuals, known as parents, for generating offspring. Among various techniques, fitness proportionate and tournament selection stand out as common methods. Fitness proportionate selection requires more computational effort than tournament selection because it involves extra calculations for each individual in the population [61]. This involves computing adjusted and normalized fitness values. Tournament selection, on the other hand, which employed in this study, relies on the tournament size [62]. It starts by randomly selecting individuals from the population to form a subset equal in size to the tournament size. Then, the fitness of each individual in this subset is assessed, and the fittest individual is chosen as the parent for producing offspring. The tournament size influences the selection pressure; smaller sizes result in less pressure, while larger sizes heighten pressure, potentially leading to premature convergence to local optima, creating an elitist GP algorithm. Despite this risk, tournament selection is widely favored because it offers adjustable selection pressure. With careful selection of the tournament size, it can maintain diversity and guide the GP algorithm towards solutions without the need for additional computations involved in calculating adjusted and normalized fitness values.

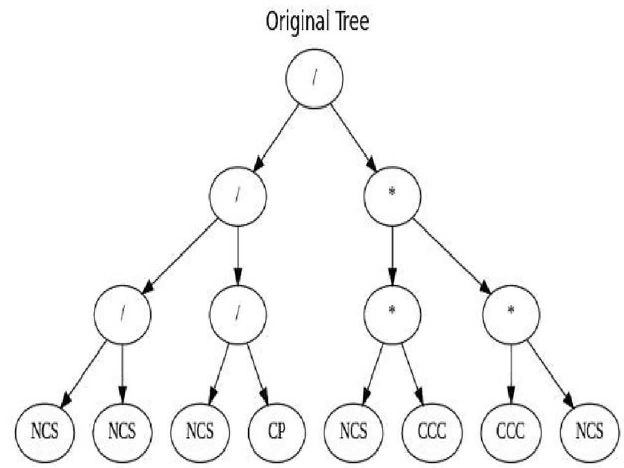

Fig. 3 Original tree

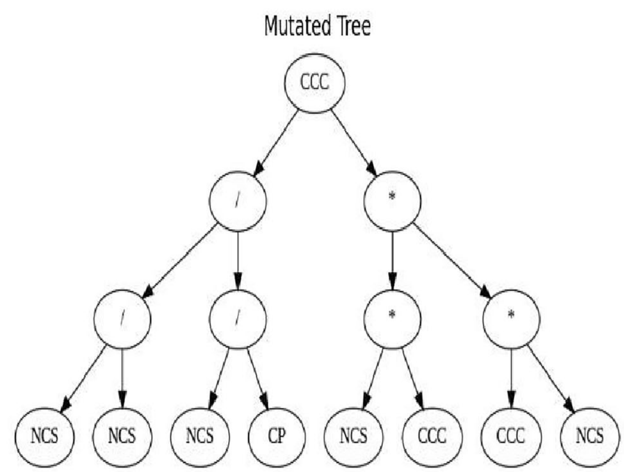

Fig. 4 Mutated tree

3.6.4 Genetic operators

Selection methods in Genetic Programming (GP) are crucial for choosing parents to generate offspring. This study employs tournament selection, where individuals are randomly chosen to form a subset of a defined tournament size. The fitness of each individual in the subset is assessed, and the fittest is selected as a parent for reproduction. The tournament size controls the selection pressure, with smaller sizes resulting in lower pressure and larger sizes potentially increasing the risk of premature convergence. However, tournament selection is preferred in this study because it allows for adjustable selection pressure, helping maintain

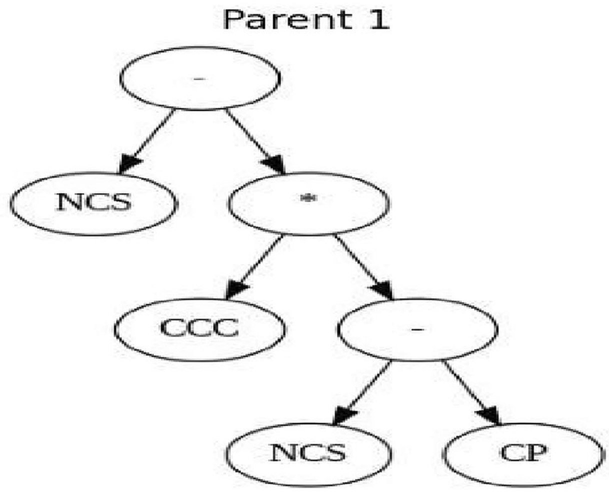

Fig. 5 Parent 1

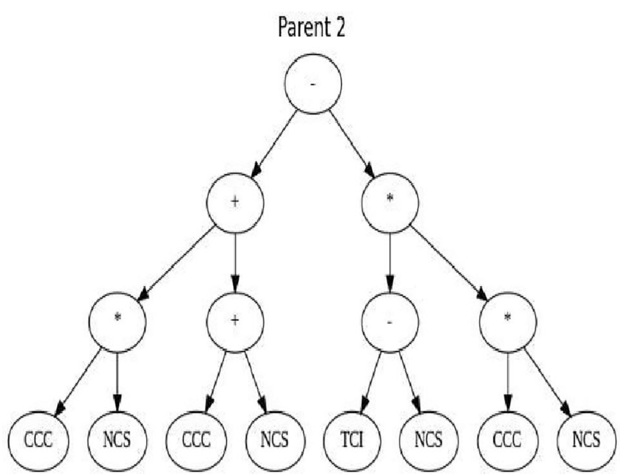

Fig. 6 Parent 2

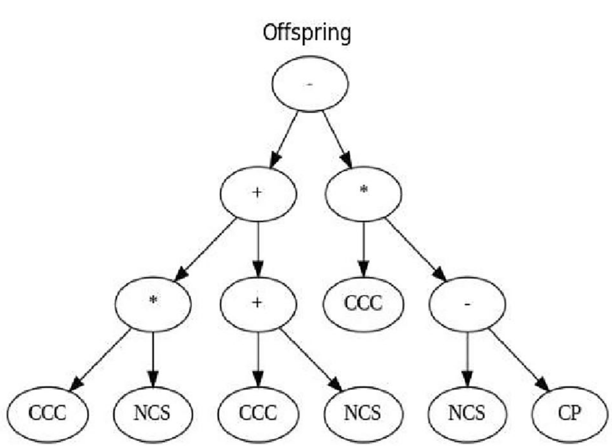

Fig. 7 New generated Offspring

diversity and guide the GP algorithm toward effective solutions.

- **Mutation**
  The mutation operator creates an offspring by altering a single parent in a specific way. First, a random mutation point, denoted as p, is selected within the parent's structure. The subtree rooted at this point is then removed, and a new subtree is randomly generated and inserted at the mutation point. To prevent excessive growth of the trees, pruning is applied. Pruning involves replacing any function node at the maximum tree depth with a terminal node. Because the mutation generates random subtrees at the mutation points, the offspring can differ significantly from the parent, making mutation a global operator. Therefore, the mutation operator does not inherently promote convergence. Figures 3 and 4 illustrates the mutation operator.

- **Crossover**
  The crossover operator, as outlined in previous studies [63, 64], produces two new offspring by merging genetic material from two parent trees. The process begins with the selection of two parents from the population using a specified selection method. Next, two crossover points, p1 and p2, are randomly chosen within the parent trees, t1 and t2. The crossover process involves extracting the subtree rooted at p1 from tree t1 and inserting it into position p2 of tree t2. Conversely, the subtree rooted at p2 is removed from t2 and inserted into position p1 of tree t1. This swapping of subtrees generates diverse offspring and aids in the convergence of the algorithm, serving as a local search mechanism, as depicted in Figures 5,6 and 7.

3.6.5 Termination and find best one

Koza describes the genetic programming (GP) approach as an ongoing process, similar to natural evolution [65]. However, in practice, the GP algorithm must conclude once a specific success criterion is achieved. This criterion is often defined as finding a solution with a 100% hit ratio, meaning a perfect solution to the problem. The definition of the success criterion can vary depending on the specific problem. In some cases, seeking a perfect solution may not be realistic, so the GP algorithm can terminate once an acceptable near-solution is found.

Algorithm 2 Comprehensive Ranking System (CRS) using Genetic Programming (GP)

Require:

Terminal Set T: {Normalized Citation Score (NCS), Temporal Citation Input (TCI), Complete Career Contribution (CCC), Collaborative Prestige (CP)}
Function Set F: {+, -, *, /}
Population Size $P = 50$
Maximum Tree Depth $D = 4$
Maximum Generations $G := 100$
Fitness Cases (Input/Output pairs)
Tournament Size $T_{\text{size}} := 5$

Ensure: Best Evolved Individual (Optimal Ranking Expression)

procedure CRS-GP
  Initialize population of size P using the ramped half-and-half method:

- Half of the trees use the full method (uniform depth).
- Half of the trees use the grow method (random shapes and sizes).

Populate each tree using T and F up to depth D.
  for all Individuals in the population **do**
    Traverse the tree to compute outputs for each fitness case.
    Sort outputs by input values.
    Compute fitness:

$$\text{Fitness} = \sum(\text{Number of fitness cases where Output = 1, in top 100}).$$

  end for
  for Generation $g = 1$ to G **do**
    Parent Selection: Use tournament selection with size T_{size}:

- Randomly select T_{size} individuals.
- Choose the individual with the highest fitness as a parent.

    Crossover:

- Select two parents using tournament selection.
- Randomly choose crossover points p_1 and p_2 in both trees.
- Swap subtrees rooted at p_1 and p_2 to create offspring.

    Mutation:

- Select a parent using tournament selection.
- Randomly choose a mutation point p in the tree.
- Replace the subtree at p with a randomly generated subtree.
- Apply pruning to limit tree depth to D.

    Replace population with offspring using survivor selection strategies.
    Evaluate fitness for the new population.
  end for
  if Stopping criterion is met (e.g., max generations or satisfactory fitness) **then**
    Terminate the algorithm.
  end if
  return Individual with the highest fitness as the optimal ranking expression.
end procedure

The genetic programming (GP) approach (represented in Algorithm 2) begins by initializing a population of random expression trees, each representing a potential ranking expression, with a maximum tree depth of 5 and evolving it over 100 generations. Each individual is evaluated using a fitness function that measures how well it ranks a set of input/output pairs, with lower rank errors corresponding to higher fitness values. Tournament selection is employed to choose parents for reproduction, where the fittest individuals are selected from random groups of 5. These parents undergo crossover and mutation operations to generate offspring with new expression trees. The offspring are evaluated, and the best individuals from both the parents and offspring are selected to form the next generation. This process continues for 100 generations, with the population gradually evolving to produce better-ranking expressions. The final result is the individual

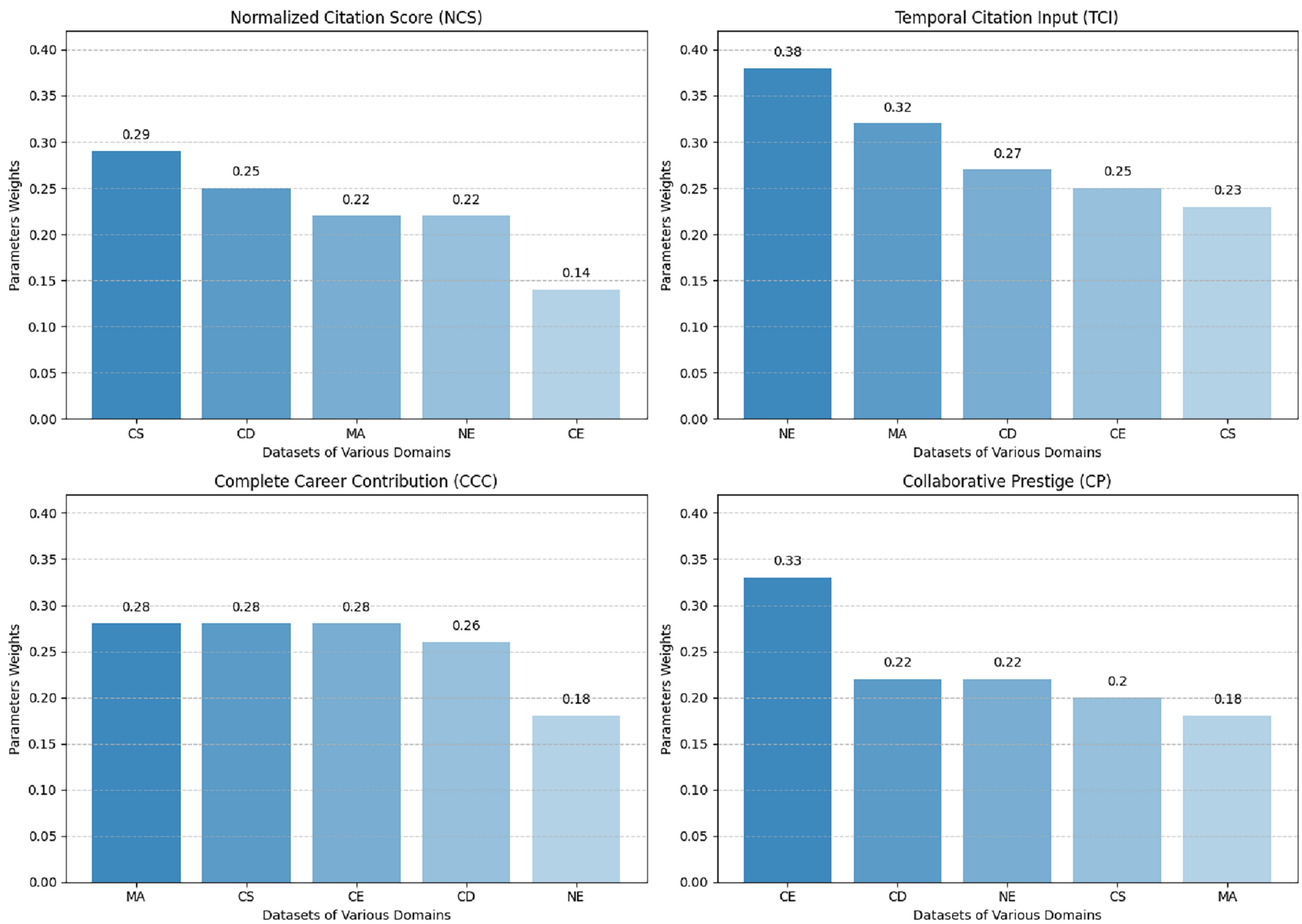

Fig. 8 Parameter weights against datasets

with the highest fitness, representing the optimal ranking expression discovered by the GP algorithm.

4 Result and discussion

The subsequent section presents the results gathered in relation to the research problem.

4.1 Parameter Weight Optimization

In this study, a multilayer neural network with a back propagation technique was applied to extract the weights of parameters, indicating their significance relative to each other based on their influence on the model's performance. The Fig. 8 presents the weights of each parameter across multiple domains.In Fig. 8, the following abbreviations are used: CS (Computer Science), MA (Mathematics), NE (Neuroscience), CE (Civil Engineering), and CD (All Fields Combined Dataset).

Analysis of the Fig. 8 reveals that, the Normalized Citation Score (NCS) is most significant in Computer Science (0.29) and the combined dataset (0.25), indicating a high value placed on citation impact in these areas. Moreover, the Temporal Citation Input (TCI) is notably highest in Neuroscience (0.38), reflecting the importance of time-sensitive citation patterns due to the rapid advancements in this field. Mathematics also shows a moderate emphasis on TCI (0.32), while Civil Engineering (0.25) and Computer Science (0.23) place less importance on it. The Complete Career Contribution (CCC) is equally valued in Civil Engineering, Computer Science, and Mathematics (0.28), highlighting the importance of career-long contributions in these fields, whereas Neuroscience (0.17) places less emphasis on cumulative career achievements. Collaborative Prestige (CP) is most crucial in Civil Engineering (0.33), underscoring the importance of collaborative efforts, likely due to the multidisciplinary nature of this domain. Neuroscience and Computer Science have moderate CP weights (0.22 and 0.19, respectively), and Mathematics has the lowest (0.18), indicating a stronger focus on individual contributions. The combined dataset reflects an intermediate level of importance for all features, balancing the diverse emphases across different fields.

4.2 Statistical ranking system (SRS)

After calculating the weight for each parameter, these weights are multiplied by their corresponding parameter values. Following this, the initial approach for proposing a new index involves using statistical methods to combine these parameters and identify the best model, which ranks the highest number of awardees in the top 100 positions. The Fig. 9 presents the result of different statistical models across multiple domains. The updated table displays the performance of various statistical models across different academic fields: Civil Engineering, Computer Science, Neuroscience, Mathematics, and a combined dataset. Each model Arithmetic Mean (AM), Harmonic Mean (HM), Contra-harmonic Mean (CHM), Geometric Mean (GM), Logarithmic Mean (LM), Root Mean Square (RMS), Trigonometric Mean (TM), and Lehmer Mean (LEM) is evaluated based on the number of awardees appearing in the top 100 records when applied to the combined features of the datasets. Computer Science continues to exhibit strong performance across most models, with consistently high values indicating significant numbers of awardees in the top 100 records. The Trigonometric Mean (TM) and Lehmer Mean (LEM) show notable improvements in several fields, particularly Mathematics, where LEM achieves the highest value of 61, indicating its effectiveness in identifying impactful records. The Logarithmic Mean (LM) remains competitive across all fields, reflecting its robustness in capturing top performers in combined datasets. Conversely, models like Arithmetic Mean (AM) and Root Mean Square (RMS) exhibit lower values across most fields, suggesting less effectiveness in this context. Overall, the Lehmer Mean (LEM) model demonstrates its superiority in the Mathematics and Computer Science datasets. The Geometric Mean proves most effective for the Neuroscience domain, while the Logarithmic Mean (LM) emerges as optimal for the Civil Engineering and combined datasets.

4.3 Comprehensive ranking system (CRS) using genetic programming (GP)

In the second approach, a genetic programming technique was employed to propose a new and comprehensive index. This technique takes parameters (Normalized Citation Score (NCS), Temporal Citation Input (TCI), Complete Career Contribution (CCC) and Collaborative Prestige (CP)) as inputs and applies all the steps of genetic programming, ultimately providing the best individual tree. In this study, four different best-fit trees were achieved across multiple domains (Fig. 9). From these best individual trees, the equation representing the combination of the proposed parameters was extracted, as presented below:

4.3.1 Computer science

After executing the genetic programming with all its steps, the technique yielded the best-fit tree shown in Fig. 10.

Upon extracting the equation from the best-fit tree, Eq. 9 was obtained:

$$CRS_{CS} = \left(\frac{CCC * CP}{CCC + TCI} \right) * (NCS * CP) \tag{9}$$

The fitness value of this best-fit tree is 0.96, indicating a high level of accuracy in capturing the relationship between the input features and the desired output. This result demonstrates the effectiveness of the genetic programming in modeling complex relationships within the dataset.

4.3.2 Civil engineering

In this section, the outcomes of applying the genetic programming technique to the civil engineering domain dataset are presented. Following the execution of the genetic programming with all its steps, the technique produced the best-fit tree shown in Fig. 11. The Eq. 10 presented the extracted equation from the best-fit tree.

$$CRS_{CE} = \left(\left((CP - TCI) * \left(\frac{CCC}{CP} \right) \right) + \left(\left(\frac{CCC}{TCI} \right) - NCS \right) \right) \tag{10}$$

The fitness value of this best-fit tree is 0.7, indicating a moderate level of accuracy in capturing the relationship between the input features and the desired output. This result highlights the potential of the genetic programming in modeling the complex interactions within the civil engineering dataset.

4.3.3 Neuroscience

In the results section for the neuroscience domain, the outcomes of applying the genetic programming technique to the dataset are presented. After executing the genetic programming with all its steps, the technique produced the best-fit tree shown in Fig. 12. The Eq. 11 presented the extracted equation from the best-fit tree.

$$CRS_{NU} = \left(\frac{NCS + CCC}{CCC} \right) - \left(\frac{CCC}{NCS} + (CP + NCS) \right) \tag{11}$$

The fitness value of this best-fit tree is 0.72, indicating a reasonably good accuracy in capturing the relationship between the input features and the desired output. This result demonstrates the efficacy of the genetic programming in modeling the complex interactions within the neuroscience dataset.

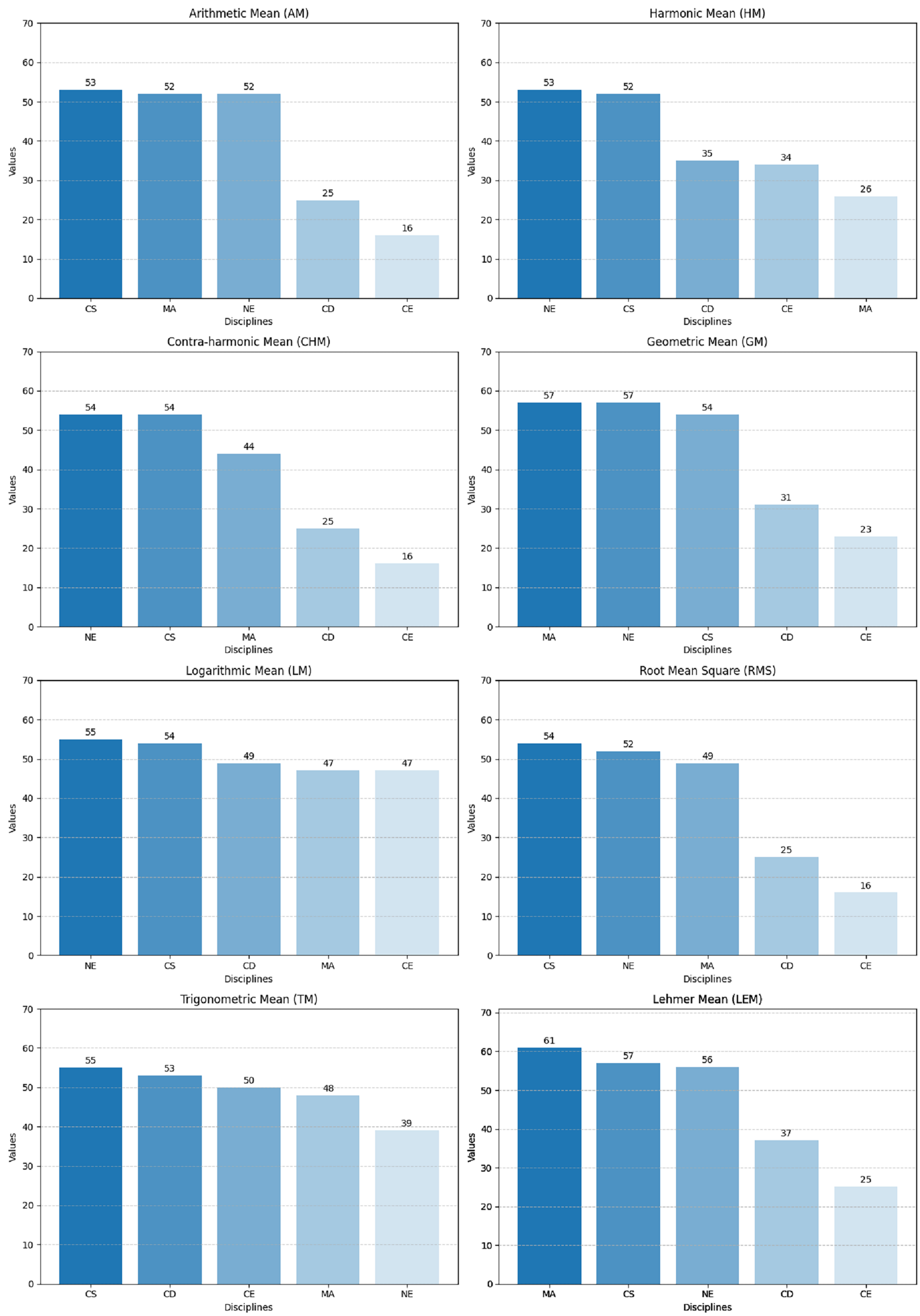

Fig. 9 Percentage of awardees agains statistical models

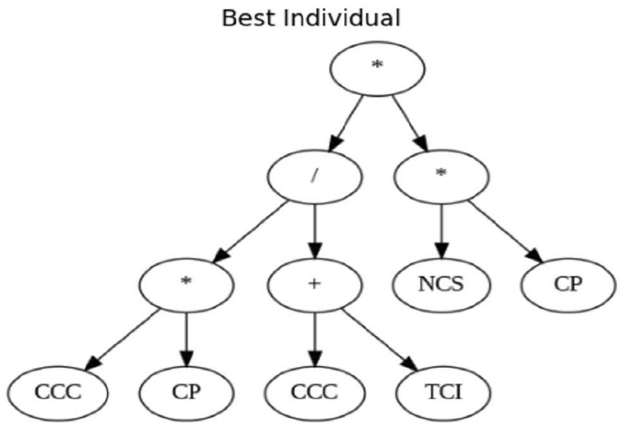

Fig. 10 Best Individual tree against computer science dataset

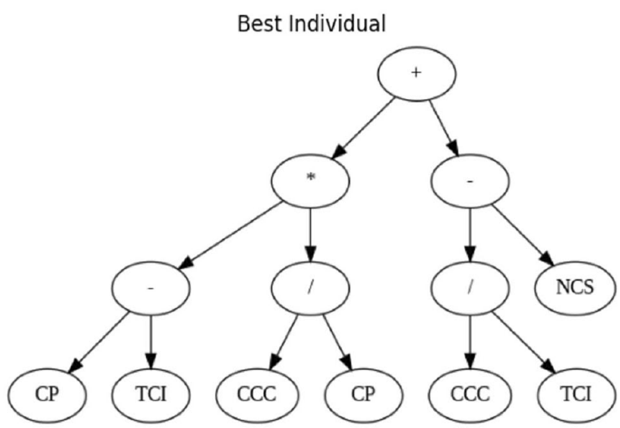

Fig. 11 Best Individual tree against civil engineering dataset

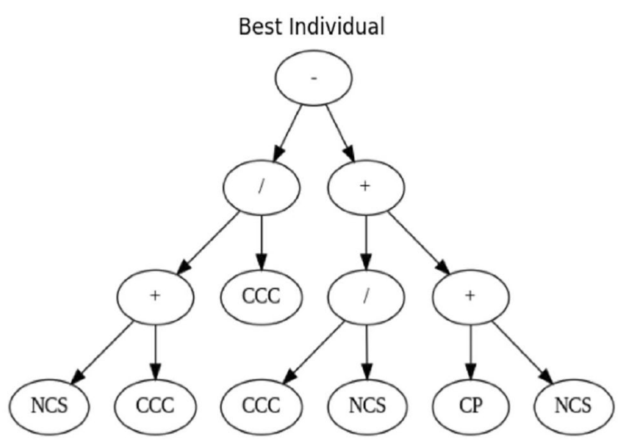

Fig. 12 Best Individual tree against Neuroscience Dataset

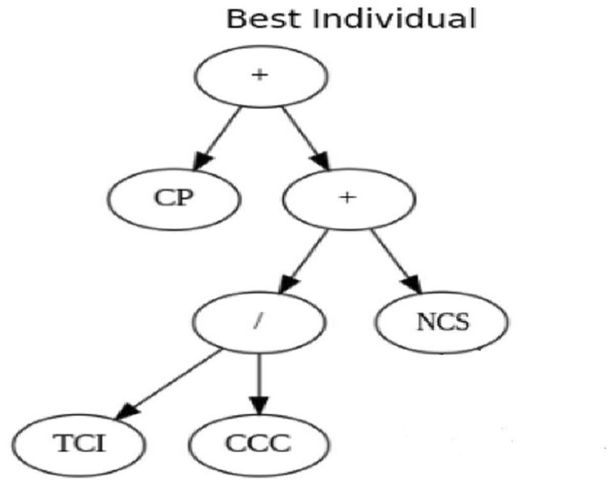

Fig. 13 Best Individual tree against mathematics dataset

4.3.4 Mathematics

In this section, the outcomes of applying the genetic programming technique to the mathematics domain dataset are presented. After executing the genetic programming with all its steps, the technique produced the best-fit tree shown in Fig. 13. The Eq. 12 presented the extracted equation from the best-fit tree.

$$CRS_{MA} = CP + \left(\frac{TCI}{CCC}\right) + NCS \tag{12}$$

The fitness value of this best-fit tree is 0.88, indicating a high level of accuracy in capturing the relationship between the input features and the desired output. This result highlights the effectiveness of the genetic programming in modeling the complex interactions within the mathematics dataset.

4.3.5 Combine all dataset

In the results section for the combined dataset, the outcomes of applying the genetic programming technique across multiple domains, including computer science, civil engineering, neuroscience, and mathematics, are presented in Fig. 14. After executing the genetic programming with all its steps, the technique produced the best-fit tree for the combined dataset. The Eq. 13 presented the extracted equation from the best-fit tree.

$$CRS_G = (((CP * TCI) + NCS) - ((TCI - NCS) * (TCI + CCC))) \tag{13}$$

The fitness value of this best-fit tree is 0.66, indicating a moderate level of accuracy in capturing the relationships between the input features and the desired output across the combined dataset. This result demonstrates the genetic

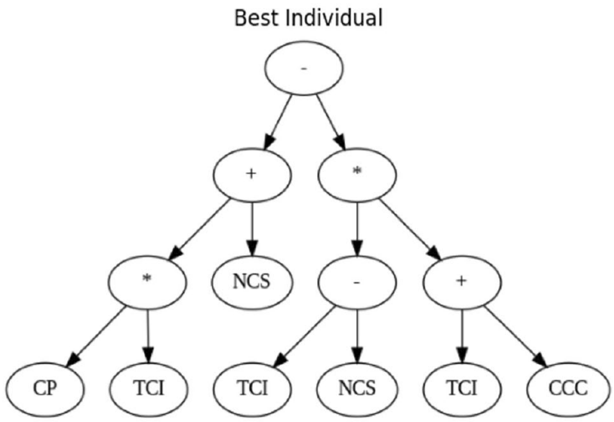

Fig. 14 Best Individual tree against combine datasets

programming capability to model complex interactions within a diverse and integrated dataset.

4.3.6 Comparative analysis

- The computer science domain achieved the highest fitness value (0.96), showcasing the genetic programming exceptional capability to model relationships in this specific field.
- The mathematics domain followed with a high fitness value of 0.88, indicating a strong and accurate model fit.
- The neuroscience domain produced a fitness value of 0.72, and the civil engineering domain had a fitness value of 0.7, both reflecting moderate accuracy.
- The combined dataset had the lowest fitness value (0.66), which is expected given the increased complexity and diversity of the integrated data.

Overall, the genetic programming demonstrated varying levels of effectiveness across different domains, with the highest accuracy observed in the computer science and mathematics domains. The results highlight the genetic programming robustness and flexibility in uncovering complex relationships within and across various fields of study.

4.4 Comparison with existing studies

For the comparison of our proposed ranking system across all domains, we conducted a literature review to identify relevant studies that have already been conducted in this area. After identifying these studies, we compared our proposed results with the findings from these studies, as shown in Fig. 15.

In the mathematics domain, Fig. 15 shows that the results of the proposed method are similar to those of the existing study [44]. However, the genetic approach is preferred because it provides a more comprehensive evaluation, covering the entire career of a researcher, whereas the existing GK index relies solely on publication and citation counts. In the computer science domain, the genetic approach demonstrates significant improvements over the existing technique [66]. In contrast, the improvement in civil engineering is minimal compared to the existing technique [45]. Similarly, in neuroscience, the genetic approach shows substantial improvement over the existing technique [39]. This domain-specific

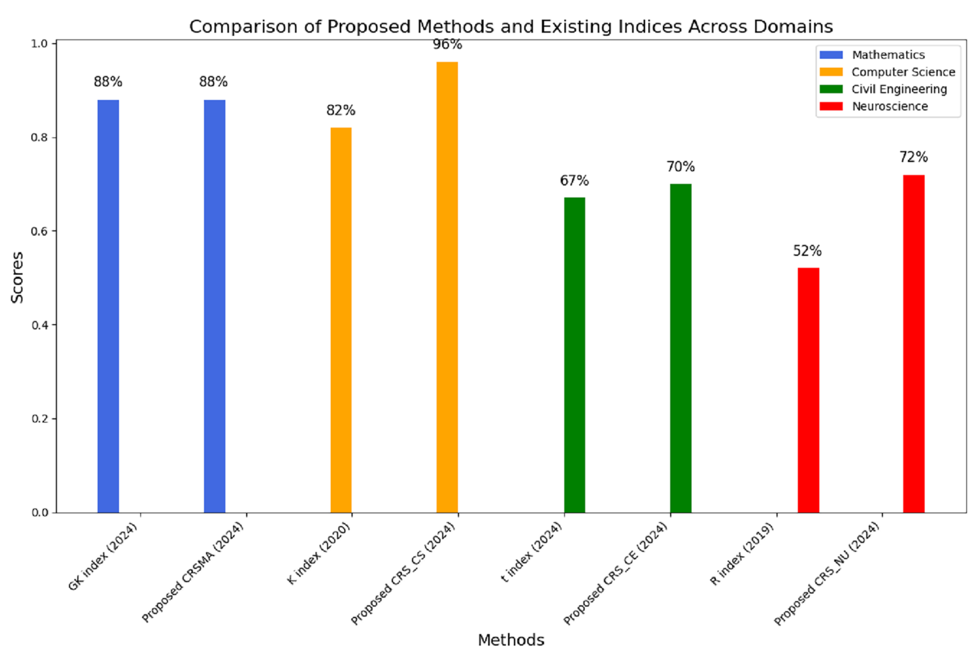

Fig. 15 Comparison with existing studies

analysis highlights that while some existing approaches yield similar results, genetic programming often outperforms them, particularly in the computer science domain, where it offers more tailored and effective solutions. Researchers in computer science frequently rely on equations generated by genetic programming. In mathematics, those seeking a fair and comprehensive evaluation may prefer the genetic-based approach, although existing statistical techniques from established indices can also be applied. In civil engineering, the genetic-based approach is more advantageous, while in neuroscience, the genetic approach shows significant benefits, making it a reliable option for researchers.

5 Discussion

The results of this study demonstrate the effectiveness of the proposed researcher ranking framework in providing a more holistic and domain-sensitive evaluation of research impact. Traditional ranking indices, such as the h-index and its variants, have long been criticized for their inability to capture qualitative aspects of research contributions. Our approach addresses these limitations by incorporating four distinct parameters-Normalized Citation Score (NCS), Complete Career Contribution (CCC), Temporal Citation Input (TCI), and Collaborative Prestige (CP)-each offering a unique perspective on a researcher's influence.

A key observation from our findings is the variation in parameter weight distribution across different domains, indicating the diverse ways in which research impact is measured. The Normalized Citation Score (NCS) received a weight of 0.29 in Computer Science and 0.25 in the combined dataset, reflecting the strong reliance on citation-based evaluation in these datasets. Similarly, the Temporal Citation Input (TCI) was assigned the highest weight (0.38) in Neuroscience, which aligns with the rapid evolution and time-sensitive nature of research in this field. Complete Career Contribution (CCC) had a weight of 0.28 in Mathematics, Civil Engineering, and Computer Science, suggesting that career-long scholarly output is a crucial factor in these domains. Finally, Collaborative Prestige (CP) was assigned a weight of 0.33 in Civil Engineering, emphasizing the role of collaboration networks in shaping research influence in this field. These weight distributions highlight how different domains emphasize various aspects of researcher impact, reinforcing the need for customized ranking models that adjust to field-specific evaluation criteria rather than relying on a one-size-fits-all approach.

Our Statistical Ranking System (SRS) further validated these observations, with models like the Lehmer Mean (LEM) and Logarithmic Mean (LM) outperforming others in identifying awardees across different fields. The effectiveness of these statistical approaches confirms that certain mathematical models are better suited for ranking researchers based on specific domain attributes. In contrast, the Comprehensive Ranking System (CRS), which leverages Genetic Programming (GP), provided a more adaptive ranking model, optimizing the weightage of different parameters based on real-world data. CRS demonstrated superior performance, achieving fitness values of 0.96 for Computer Science and 0.88 for Mathematics, showcasing its potential as a dynamic, domain-specific ranking mechanism.

Additionally, our results indicate a trade-off between domain specificity and generalization. While domain-specific models achieved higher accuracy, merging datasets resulted in a lower fitness value (0.66 for the combined dataset). This suggests that a single ranking model may not be optimal across all disciplines, reinforcing the need for customized ranking mechanisms for different research domains.

Despite these promising findings, this study has certain limitations. The proposed framework was validated on four scientific domains (Computer Science, Mathematics, Neuroscience, and Civil Engineering). Future work should explore additional fields, such as medical sciences and physics, to assess the adaptability of our approach. Moreover, while Genetic Programming (GP) proved effective in optimizing ranking models, further research can explore alternative metaheuristic optimization techniques (e.g., Potter Optimization Algorithm (POA), Carpet Weaving Optimization (CWO), and Fossa Optimization Algorithm (FOA)) to enhance ranking performance.

Overall, our study underscores the importance of moving beyond citation-based metrics and incorporating qualitative measures to assess research impact more effectively. The results highlight that research evaluation must be domain-sensitive, adaptable, and multidimensional, ensuring a fairer and more accurate ranking of researchers across diverse scientific fields.

6 Conclusion and future work

Ranking researchers in a fair and meaningful way remains a complex but essential challenge for the scientific community. Traditional metrics, such as publication and citation counts while widely used, offer only a limited perspective on a researcher's overall impact. Although composite indices like the h-index have gained popularity, they also fall short in addressing the qualitative and longitudinal aspects of scholarly contributions. This study proposed a novel approach for researcher evaluation by introducing four parameters, Normalized Citation Score (NCS), Complete Career Contribution (CCC), Temporal Citation Input (TCI), and Collaborative Prestige (CP), which collectively aim to capture both the quantitative and qualitative dimensions of academic impact. These parameters were integrated into two ranking systems: the Statistical Ranking System (SRS) and the Comprehensive Ranking System (CRS) based on Genetic Programming (GP). The evaluation across datasets from Computer Science, Mathematics, Neuroscience, Civil Engineering, and a combined dataset revealed that different parameters hold varying significance across fields. SRS demonstrated the utility of statistical models such as the Lehmer Mean and Logarithmic Mean, while CRS provided a more adaptive mechanism by evolving domain-specific equations that achieved high fitness scores in some domains. These findings suggest that ranking systems benefit from being domain-sensitive rather than domain-agnostic. The performance drop observed in the combined dataset further underscores the importance of domain-specific modeling in researcher evaluation.

While the results are encouraging, further improvements and validations are necessary. Future work will involve testing the framework in additional domains (e.g., medical sciences, physics), incorporating broader researcher profiles, and exploring alternative metaheuristic approaches (e.g., POA, FOA) to enhance the adaptability and robustness of the ranking methodology.

7 Limitations of the study

While the proposed ranking framework demonstrates promising results across several scientific domains, a few limitations should be acknowledged. First, the study is limited to four domains-Computer Science, Mathematics, Neuroscience, and Civil Engineering-which, although diverse, may not fully represent the complexities of all research fields. The results may not generalize to domains such as medical sciences or social sciences without further validation. Second, the fitness values used to assess the model performance are derived from awardee-based ground truths, which may reflect implicit biases in award distribution practices. Lastly, although Genetic Programming (GP) offers flexibility in evolving domain-specific ranking formulas, its computational overhead and sensitivity to parameter settings may limit scalability to larger datasets or real-time applications.

Acknowledgements This research did not receive funding from any organization. Muhammad Tanvir Afzal conceived the idea and supervised the research, Ghulam Mustafa and Muhammad Abdullah conducted the experiments, analyzed the results, and wrote the paper, while Abid Rauf reviewed the paper and supervised the research.

Author Contributions Muhammad Tanvir Afzal conceived the idea and supervised the research. Ghulam Mustafa and Muhammad Abdullah conducted the experiments, analyzed the results, and wrote the paper. Abid Rauf reviewed the paper and supervised the research.

Funding This research did not receive funding from any organization.

Data Availability No datasets were generated or analysed during the current study.

Declarations

Competing interests The authors declare no competing interests.

References

1. Ahmed B, Li W, Mustafa G, Afzal MT, Alharthi SZ, Akhunzada A (2023) Evaluating the effectiveness of author-count based metrics in measuring scientific contributions. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3309416
2. Mustafa G, Usman M, Yu L, Afzal MT, Sulaiman M, Shahid A (2021) Multi-label classification of research articles using word2vec and identification of similarity threshold. Sci Rep 11(1):21900
3. Xia W, Li T, Li C (2023) A review of scientific impact prediction: tasks, features and methods. Scientometrics 128(1):543–585. https://doi.org/10.1007/s11192-022-04547-8
4. Jiang X, Sun X, Zhuge H (2013) Graph-based algorithms for ranking researchers: not all swans are white! Scientometrics 96:743–759. https://doi.org/10.1007/s11192-012-0943-y
5. Shah SMAH, Ullah A, Iqbal J, Bourouis S, Ullah SS, Hussain S, Khan MQ, Shah YA, Mustafa G (2023) Classifying and localizing abnormalities in brain mri using channel attention based semi-bayesian ensemble voting mechanism and convolutional auto-encoder. IEEE Access 11:75528–75545
6. Raheel M, Ayaz S, Afzal MT (2018) Evaluation of h-index, its variants and extensions based on publication age & citation intensity in civil engineering. Scientometrics 114:1107–1127. https://doi.org/10.1007/s11192-017-2633-2
7. Mustafa G, Usman M, Afzal MT, Shahid A, Koubaa A (2021) A comprehensive evaluation of metadata-based features to classify research paper's topics. IEEE Access 9:133500–133509
8. Bihari A, Tripathi S, Deepak A (2023) A review on h-index and its alternative indices. J Inf Sci 49(3):624–665. https://doi.org/10.1177/01655515211014478
9. Mustafa G, Rauf A, Al-Shamayleh AS, Sulaiman M, Alrawagfeh W, Afzal MT, Akhunzada A (2023) Optimizing document classification: unleashing the power of genetic algorithms. IEEE Access 11:83136–83149
10. Mustafa G, Rauf A, Al-Shamayleh AS, Ahmed B, Alrawagfeh W, Afzal MT, Akhunzada A (2023) Exploring the significance of publication-age-based parameters for evaluating researcher impact. IEEE Access. https://doi.org/10.1109/ACCESS.2023.3304013
11. Ahmed B, Wang L, Al-Shamayleh AS, Afzal MT, Mustafa G, Alrawagfeh W, Akhunzada A (2023) Machine learning approach for effective ranking of researcher assessment parameters. IEEE Access 11:133294–133312
12. Prathap G (2010) The 100 most prolific economists using the p-index. Scientometrics 84(1):167–172. https://doi.org/10.1007/s11192-009-0068-0
13. Ahmed B, Wang L, Hussain W, Mustafa G, Afzal MT (2025) Investigating scholarly indices and their contribution to recognition patterns among awarded and non-awarded researchers. Int J Data Sci Anal, pp 1–18

14. Adnan SM, Ahmad W, Mahmood I, Mustafa G, Dattana V et al (2024) Enhancing text mining efficiency using an effective topic modeling approach. Tech J 29(01):39–46
15. Alshdadi AA, Usman M, Alassafi MO, Afzal MT, AlGhamdi R (2023) Formulation of rules for the scientific community using deep learning. Scientometrics 128(3):1825–1852. https://doi.org/10.1007/s11192-023-04633-5
16. Liu J-X, Yin M-M, Gao Y-L, Shang J, Zheng C-H (2022) Msf-lrr: multi-similarity information fusion through low-rank representation to predict disease-associated microbes. IEEE/ACM Trans Comput Biol Bioinf 20(1):534–543. https://doi.org/10.1109/TCBB.2022.3146176
17. Kanwal A, Masood N, Mustafa G, Ghafoor MA, Ayaz S (2025) Mk-smote and m-smote: enhanced techniques for handling class imbalance problem. Iran J Comput Sci, pp 1–19
18. Mustafa G, Rauf A, Ahmed B, Afzal MT, Akhunzada A, Alharthi SZ (2023) Comprehensive evaluation of publication and citation metrics for quantifying scholarly influence. IEEE Access 11:65759–65774. https://doi.org/10.1109/ACCESS.2023.3290917
19. Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci 102(46):16569–16572. https://doi.org/10.1073/pnas.050765510
20. Chen Z, Yao J, Xiao G, Wang S (2021) Efficient and differentiable low-rank matrix completion with back propagation. IEEE Trans Multimed. https://doi.org/10.1109/TMM.2021.3124087
21. Tol R (2009) The h-index and its alternatives: an application to the 100 most prolific economists. Scientometrics 80(2):317–324. https://doi.org/10.1007/s11192-008-2079-7
22. Jin B, Liang L, Rousseau R, Egghe L (2007) The r-and ar-indices: complementing the h-index. Chin Sci Bull 52(6):855–863. https://doi.org/10.1007/s11434-007-0145-9
23. Zhang C-T (2009) The e-index, complementing the h-index for excess citations. PLoS ONE 4(5):5429. https://doi.org/10.1371/journal.pone.0005429
24. Aziz NA, Rozing MP (2013) Profit (p)-index: the degree to which authors profit from co-authors. PLoS ONE 8(4):59814. https://doi.org/10.1371/journal.pone.0059814
25. Burrell QL (2007) On the h-index, the size of the hirsch core and jin's a-index. J Informet 1(2):170–177. https://doi.org/10.1016/j.joi.2007.01.003
26. Lopez J, Susarla SM, Swanson EW, Calotta N, Lifchez SD (2015) The association of the h-index and academic rank among full-time academic hand surgeons affiliated with fellowship programs. J Hand Surg 40(7):1434–1441. https://doi.org/10.1016/j.jhsa.2015.03.026
27. Kosmulski M et al (2006) A new hirsch-type index saves time and works equally well as the original h-index. ISSI Newsletter 2(3):4–6. https://doi.org/10.1177/01655515211014478
28. Cabrerizo FJ, Alonso S, Herrera-Viedma E, Herrera F (2010) q2-index: quantitative and qualitative evaluation based on the number and impact of papers in the hirsch core. J Informet 4(1):23–28. https://doi.org/10.1016/j.joi.2009.06.005
29. Khan NR, Thompson CJ, Taylor DR, Gabrick KS, Choudhri AF, Boop FR, Klimo P Jr (2013) Part ii: Should the h-index be modified? an analysis of the m-quotient, contemporary h-index, authorship value, and impact factor. World Neurosurg 80(6):766–774. https://doi.org/10.1016/j.wneu.2013.07.011
30. Salman M, Ahmed MM, Afzal MT (2021) Assessment of author ranking indices based on multi-authorship. Scientometrics 126(5):4153–4172. https://doi.org/10.1007/s11192-021-03906-1
31. Katsaros D, Akritidis L, Bozanis P (2009) The f index: quantifying the impact of coterminal citations on scientists' ranking. J Am Soc Inform Sci Technol 60(5):1051–1056. https://doi.org/10.1002/asi.21040
32. Cameron DHL, Aleman-Meza B, Decker S, Arpinar IB (2007) Semef: A taxonomy-based discovery of experts, expertise and collaboration networks. PhD thesis, University of Georgia. https://doi.org/10.1177/01655515211014478
33. Ye F, Rousseau R (2010) Probing the h-core: an investigation of the tail-core ratio for rank distributions. Scientometrics 84(2):431–439. https://doi.org/10.1007/s11192-009-0099-6
34. Van Raan AF (2006) Comparison of the hirsch-index with standard bibliometric indicators and with peer judgment for 147 chemistry research groups. Scientometrics 67:491–502. https://doi.org/10.1556/Scient.67.2006.3.10
35. Schreiber M (2007) Self-citation corrections for the hirsch index. Europhys Lett 78(3):30002. https://doi.org/10.1209/0295-5075/78/30002
36. Xiao S, Yan J, Li C, Jin B, Wang X, Yang X, Chu SM, Zha H (2016) On modeling and predicting individual paper citation count over time. In: Ijcai, pp. 2676–2682. https://doi.org/10.1109/SCEECS57921.2023.10061818
37. Ayaz S, Afzal MT (2016) Identification of conversion factor for completing-h index for the field of mathematics. Scientometrics 109(3):1511–1524. https://doi.org/10.1007/s11192-016-2122-z
38. Ghani R, Qayyum F, Afzal MT, Maurer H (2019) Comprehensive evaluation of h-index and its extensions in the domain of mathematics. Scientometrics 118:809–822. https://doi.org/10.1007/s11192-019-03007-0
39. Ameer M, Afzal MT (2019) Evaluation of h-index and its qualitative and quantitative variants in neuroscience. Scientometrics 121(2):653–673. https://doi.org/10.1007/s11192-019-03209-6
40. Ain Q-U, Riaz H, Afzal MT (2019) Evaluation of h-index and its citation intensity based variants in the field of mathematics. Scientometrics 119:187–211. https://doi.org/10.1007/s11192-019-03009-y
41. Usman M, Mustafa G, Afzal MT (2021) Ranking of author assessment parameters using logistic regression. Scientometrics 126(1):335–353. https://doi.org/10.1007/s11192-020-03769-y
42. Mustafa G, Rauf A, Tanvir Afzal M (2024) Mret: modified recursive elimination technique for ranking author assessment parameters. PLoS ONE 19(6):0303105. https://doi.org/10.1371/journal.pone.0303105
43. Mustafa G, Rauf A, Al-Shamayleh AS, Afzal MT, Waqas A, Akhunzada A (2024) Defining quantitative rules for identifying influential researchers: insights from mathematics domain. Heliyon. https://doi.org/10.1016/j.heliyon.2024.e30318
44. Mustafa G, Rauf A, Afzal MT (2024) Gk index: bridging gf and k indices for comprehensive author evaluation. Knowl Inform Syst. https://doi.org/10.1007/s10115-024-02119-1
45. Mustafa G, Rauf A, Afzal MT (2024) Enhancing author assessment: an advanced modified recursive elimination technique (mret) for ranking key parameters and conducting statistical analysis of top-ranked parameter. Int J Data Sci Anal. https://doi.org/10.1007/s41060-024-00545-6
46. Hamadneh T, Batiha B, Alsayyed O, Bektemyssova G, Montazeri Z, Dehghani M, Eguchi K (2024) On the application of potter optimization algorithm for solving supply chain management application. Int J Intell Eng Syst 17(5)
47. Alomari S, Kaabneh K, AbuFalahah I, Gochhait S, Leonova I, Montazeri Z, Dehghani M, Eguchi K (2024) Carpet weaver optimization: a novel simple and effective human-inspired metaheuristic algorithm. Int J Intell Eng Syst 17(4)
48. Hamadneh T, Batiha B, Werner F, Montazeri Z, Dehghani M, Bektemyssova G, Eguchi K (2024) Fossa optimization algorithm: a new bio-inspired metaheuristic algorithm for engineering applications. Int J Intell Eng Syst 17(5):1038–1047
49. Hamadneh T, Kaabneh K, Alssayed O, Eguchi K, Gochhait S, Leonova I, Dehghani M (2024) Addax optimization algorithm: a novel nature-inspired optimizer for solving engineering

applications. Int J Intell Eng Syst. https://doi.org/10.22266/ijies2024.0630.57
50. Kaabneh K, AbuFalahah I, Eguchi K, Gochhait S, Leonova I, Montazeri Z, Dehghani M et al (2024) Dollmaker optimization algorithm: a novel human-inspired optimizer for solving optimization problems. Int J Intell Eng Syst. https://doi.org/10.22266/ijies2024.0630.63
51. Hamadneh T, Kaabneh K, AlSayed O, Bektemyssova G, Montazeri Z, Dehghani M, Eguchi K (2024) Sculptor optimization algorithm: a new human-inspired metaheuristic algorithm for solving optimization problems. Int J Intell Eng Syst. https://doi.org/10.22266/ijies2024.0831.43
52. Hamadneh T, Batiha B, Al-Baik O, Bektemyssova G, Montazeri Z, Werner F, Dhiman G, Dehghani M, Eguchi K (2024) Sales training based optimization: A new human-inspired metaheuristic approach for supply chain management. International Journal of Intelligent Engineering & Systems 17(6)
53. Hamadneh T, Batiha B, Gharib GM, Montazeri Z, Werner F, Dhiman G, Dehghani M, Jawad RK, Aram E, Ibraheem IK et al (2025) Orangutan optimization algorithm: an innovative bio-inspired metaheuristic approach for solving engineering optimization problems. Int J Intell Eng Syst 18(1):45–58
54. Hamadneh T, Batiha B, Alsayyed O, Montazeri Z, Ashtiani HJ, Jafarpour M, Dehghani M, Eguchi K (2025) On the application of tailor optimization algorithm for solving real-world optimization application. Int J Intell Eng Syst 18(1)
55. Hamadneh T, Batiha B, Al-Baik O, Montazeri Z, Malik OP, Werner F, Dhiman G, Dehghani M, Eguchi K (2025) Spider-tailed horned viper optimization: An effective bio-inspired metaheuristic algorithm for solving engineering applications. Int J Intell Eng Syst 18(1)
56. Black JE, Kueper JK, Williamson TS (2023) An introduction to machine learning for classification and prediction. Fam Pract 40(1):200–204. https://doi.org/10.1093/fampra/cmac104
57. Huang J, Zhang J, Li X, Qiao Y, Zhang R, Kumar GS (2023) Investigating the effects of ensemble and weight optimization approaches on neural networks' performance to estimate the dynamic modulus of asphalt concrete. Road Mater Pavement Des 24(8):1939–1959. https://doi.org/10.1080/14680629.2022.2112061
58. Zeng X, Martinez TR (2000) Using a neural network to approximate an ensemble of classifiers. Neural Process Lett 12:225–237. https://doi.org/10.1023/A:1026530200837
59. Zhang F, Mei Y, Nguyen S, Zhang M (2023) Survey on genetic programming and machine learning techniques for heuristic design in job shop scheduling. IEEE Trans Evol Comput 28(1):147–167. https://doi.org/10.1023/10.1109/TEVC.2023.3255246
60. Vanneschi L, Silva S (2023) Genetic programming. In: Lectures on Intelligent Systems, pp. 205–257. https://doi.org/10.1109/TEVC.2023.3255246
61. Farinati D, Bakurov I, Vanneschi L (2023) A study of dynamic populations in geometric semantic genetic programming. Inf Sci 648:119513. https://doi.org/10.1016/j.ins.2023.119513
62. Xu M, Mei Y, Zhang F, Zhang M (2023) Genetic programming with lexicase selection for large-scale dynamic flexible job shop scheduling. IEEE Trans Evol Comput. https://doi.org/10.1109/TEVC.2023.3244607
63. Pretorius K, Pillay N (2024) Neural network crossover in genetic algorithms using genetic programming. Genet Program Evolvable Mach 25(1):7. https://doi.org/10.1007/s10710-024-09481-7
64. Fan Q, Bi Y, Xue B, Zhang M (2024) A genetic programming-based method for image classification with small training data. Knowl-Based Syst 283:111188. https://doi.org/10.1016/j.knosys.2023.111188
65. Koza JR (2010) Human-competitive results produced by genetic programming. Genet Program Evolvable Mach 11:251–284. https://doi.org/10.1007/s10710-010-9112-3
66. Ayaz S, Masood N (2020) Comparison of researchers' impact indices. PLoS ONE 15(5):0233765