Knowledge and Information Systems (2025) 67:9987–10004
https://doi.org/10.1007/s10115-025-02522-2

RESEARCH

Enhancing researcher evaluation in computer science: a novel index for impact assessment

Hafiza Zarafshan Mukhtiar[1] · Ghulam Mustafa[1] · Muhammad Tanvir Afzal[1]

Received: 18 March 2025 / Revised: 12 May 2025 / Accepted: 14 June 2025 /
Published online: 29 June 2025

Abstract
Evaluating researcher impact is a critical aspect of scientometrics, influencing key academic decisions such as promotions, funding allocations, and prestigious awards. While traditional bibliometric indicators such as the h-index and its various extensions are widely adopted, ongoing debates persist regarding their accuracy in capturing true research impact. This study introduces a novel index aimed at improving researcher evaluation in the field of computer science. Using a dataset of 1200 researchers, evenly split between 600 award recipients and 600 non-recipients, we systematically evaluate individual bibliometric indices to identify the most effective parameters. These top-performing parameters are then combined using advanced statistical models including the geometric mean, harmonic mean, and contra-harmonic mean, etc., to determine the optimal aggregation method for impact assessment. The outcome of this study indicate that the h2-upper index and the AWCR index demonstrate the highest predictive accuracy among individual indicators. Moreover, the contra-harmonic mean proves to be the most effective statistical model for combining parameters. Based on these insights, we propose a new index derived by applying the contra-harmonic mean to the best-performing parameter pair. This enhanced framework offers a more robust and reliable approach to researcher evaluation, supporting improved academic assessment and recognition in computer science.

Keywords Author impact assessment · Statistical models · H index · Parameter ranking · Scientometrics

1 Introduction

Evaluation of scientific output in the contemporary research environment is important, yet very challenging [1–4]. Dorta-Gonzalez and Dorta-González (2013) [5] report that a clearly defined hierarchy significantly aids in identifying exceptional peers across disciplines and

✉ Hafiza Zarafshan Mukhtiar
zarafshan377@gmail.com

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

Muhammad Tanvir Afzal
director.ssc@stmu.edu.pk

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

awarding them high status, resources, or academic positions. Such systems are valuable for measuring academic productivity and informing collaborative decisions. For example, Raheel et al. (2018) argue that a fair and comprehensive ranking system can help conference organizers select keynote speakers and help students choose research supervisors based on proven capabilities of researchers [6]. Evaluation of research work is highly dependent on standards established within the scientific community [7].

According to Dorta-Gonzalez and Dorta-González (2013), a comprehensive ranking system can address key questions such as identifying which researchers deserve scholarships, promotions, and recognition, as well as determining who contributes the most innovative and impactful research [5]. Various indices have been proposed to measure authors' contributions to the academic community [1]. As James (2014) notes, numerous parameters can be used to assess a researcher's importance [8]. The literature presents a wide array of measures aimed at evaluating the value of scientists, and many techniques have been developed over time to rank authors based on academic performance [9]. Whether qualitative or quantitative, these methods strive to establish fair and meaningful criteria for assessment. Nevertheless, the absence of a clear consensus on how to define and measure scholarly impact continues to hinder the development of robust evaluation systems.

In scientometric research, the concept of impact is inherently complex and multidimensional. It goes beyond mere citation counts and encompasses influence across academic, societal and technological dimensions. Waltman (2016) emphasizes that the design of bibliometric indicators should be grounded in a clear theoretical understanding of what is being measured and why, rather than relying solely on empirical robustness [10]. Similarly, Abramo and D'Angelo (2018) argue that bibliometric indicators are only meaningful if rooted in a well-defined concept of impact and supported by transparent assumptions [11].

This study operationalizes impact using prestigious international research awards (e.g., IEEE Fellowships, the Turing Award, ACM Distinctions) as proxies for peer-recognized excellence. Although this approach may not fully capture emerging subfields or informal contributions, it provides an externally validated and domain-recognized benchmark. The proposed index is not intended to universally define impact, but to optimize bibliometric assessment within this validation framework.

We acknowledge the ongoing global shift toward more holistic research evaluation models, as advocated by initiatives such as CoARA (Coalition to Advance Research Assessment) and recent critiques by Torres-Salinas et al. (2023), which highlight the limitations of reductionist metrics based solely on citations [12]. This framework, though quantitative, is designed to complement not replace these evolving qualitative approaches.

Historically, assessment systems have relied on published papers or citation counts [13]. However, there has been growing interest in developing indices that assess not only the quantity but also the influence of a researcher's work [14]. These newer metrics aim to more comprehensively capture both author productivity and research impact [15]. One of the most prominent is the h-index, introduced by Hirsch (2005), which quantifies both the quality and quantity of a researcher's output [16]. While widely adopted, the h-index has well-known limitations, such as failing to account for highly cited outlier papers or adjusting for disciplinary norms [17]. This has led to the development of numerous alternatives, such as the g-index [18], a-index [3], r-index [19], q2-index [20], and hg-index [21], each attempting to address specific gaps.

Even with these alternatives, the absence of a universally accepted standard for evaluating scientific impact persists. Many proposed methodologies rely on hypothetical case studies or narrowly focused datasets, limiting their generalizability [1, 20, 22]. Therefore, there is an urgent need for a systematic and comprehensive study that evaluates a broad range of indicators within a single, well-defined domain using an external validation mechanism.

This study addresses that need by developing a targeted index for assessing researcher impact in the domain of computer science. Using a dataset of 1,200 researchers, evenly divided between 600 award recipients and 600 non-recipients, we investigate the performance of 64 bibliometric parameters in distinguishing recognized scholarly excellence. These rankings are validated against prestigious awards from professional bodies such as IEEE, ACM, and the Turing Award. We evaluate the parameters both individually and in pairwise combinations using seven statistical models: arithmetic mean, harmonic mean, geometric mean, quadratic mean, cubic mean, contra-harmonic mean, and logarithmic mean. The final proposed index, based on the contra-harmonic mean of the h2-upper index and AWCR, performs best among the evaluated pairs.

The remainder of the paper is structured as follows: Section 2 reviews the related work and existing methodologies; Section 3 outlines the methodology; Section 4 presents the key findings; and Sect. 5 concludes with potential directions for future work.

2 Related work

In the academic context, measuring the empirical output of researchers is important in making academic choices such as awarding honors, promotions or other activities in their career progression [23]. Various measures which comprise the number of publications, citations, co-authors and an assortment of other indices, including the h-index, g-index, a-index and many others, are used to evaluate research output. However, the underlying traditional measures of focus on elementary parameters such as publications and citations profoundly limit these metrics. There is inadequate research activity for detailed individual assessment at a global scale. One major problem surfaces when trying to compare researchers with different goals or methods. One researcher might publish a huge volume of papers every year, whereas another might be a highly innovative quality researcher [24]. When researchers with different intentions are subjected to evaluation by these traditional indices, the situation is even worse. Higher publication count researchers are hypothesized to be more productive, but that is not always the case. To illustrate, an active scholar may publish in low impact journals or present at low-level conferences, which subsequently reduces the significance of their input [25]. Likewise, one's citation count, which is frequently viewed as a dependable measurement, holds some intricacies. A researcher's citations are susceptible to change because of self-serving actions like self-citation or because the citations stem from a researcher's name instead of the actual work [19]. In addition, citations depending on critique or acknowledgment do not always indicate the actual mark or relevance of the research, raising issues of concern when relying solely on citation counts.

As both a researcher's publications and citations are important, Hirsch put forth the idea of using the h-index to track such data [16]. The simplicity of the h-index is one of the most frequently cited reasons for its notoriety within the scientific community. But, like most things, the h-index carries flaws as well. For instance, once citations to a researcher's h-core papers (the most cited publications) exceed a certain threshold, they no longer contribute to increasing the h-index. Furthermore, this metric does not work well for new researchers, as it requires time for a researcher to publish and then accumulate citations. This issue becomes more problematic for early-career researchers who may be highly innovative but have yet to publish enough work to accumulate the citations needed to boost their h-index. Moreover, the h-index tends to favor inactive or senior researchers, whose prior work continues to accumulate citations, even though they may not be actively contributing to new scientific advancements [17]. To overcome the shortcomings of the h-index, several alternative indices have been proposed. Citation intensity-based indices, such as the A-index [15], R-index [19], e-index [26], and f-index, focus on the quality and strength of citations rather than just their quantity. The A-index, proposed by Rousseau [15], is based on average citations of a reviewer's Hirsch-core, which only includes the most cited papers. However, the A-index may penalize researchers with many highly cited papers because, as more high-impact papers are published, the average citation count decreases. To remedy this, the R-index was proposed as a substitute by taking the square root of the sum of citations instead of an average [19]. This is generally accepted as a better measure of a researcher's citation productivity.

Even though the A-index and R-index were created to improve upon the h-index, they still omit the contributions of papers absent from the h-core [19]. For example, let's suppose a researcher publishes ten papers. If five of these papers have five or more citations, the remaining five papers, which may hold useful information, do not contribute anything toward the h-index. To solve this problem, the e-index was formulated which encompasses all publications and citations, not only the ones in the h-core [26]. Correspondingly, the f-index incorporates co-terminal citations, introducing a new and more appropriate way of measuring the h-core papers' citations by assigning them a fraction of the total equaled to the h-core papers' fold [27]. Alongside these indices that rely on citation counts, Sidiropoulos et al. (2007) [28] created two other forms based on the h-index with a more flexible structure: the trend h-index (ht-index) and the contemporary h-index (hc-index). These indices consider trends in the rate of citing particular articles which marks an improvement upon the classical h-index. Studies indicate that these adaptations likely outperform the original h-index in terms of their capability of detecting the influential research works [28]. Wu in 2010 [29] introduced the w-index in order to shift the focus from generalized impact measures to notoriety of works and, therefore, most likely contributed to a more precise definition of a researcher's influence. This approach appears to fundamentally differ from the h-index by placing greater emphasis on the most impactful documents. Ameer performed a detailed evaluation of the h-index and its modifications as applied to neurosurgery [30], while Van Raan [31] examined h-index measurements in chemistry research teams. All of these studies indicate that there is a need for a more nuanced methodology that combines other elements besides publication or citation metrics.

Ayaz and Afzal [32] provided a new solution to the examination of h-indices in award-winning members of scientific societies in the Mathematics field in 2016. Their research suggested that when awards were used as a metric, the complete-h outperformed the g-index and h-index [32]. Moreover, studies focused on analyzing the performance of h-index and g-index alongside q2-index and other gauges on Mathematics datasets found that h-index outperformed others considering the winning researchers in focus. Nonetheless, during their analysis, Ain et al. (2019) [24] found that the h-index only captured 33 out of 5753 authors, which proved the h-index needed revising. [6] advanced the civil engineering work by citing other research through which it became evident that citation-based measures do not serve as a proxy for research quality. These researchers stated that papers with the most awards seem to be the least cited, implying that citations are not guaranteed markers of quality. Ameer and Afzal (2019) [30] studied fields like neuroscience and incorporated g-index, hg-index, m-quotient, e-index, f-index, and r-index with award-winning and their results proved that hg-index and r-index superseded other metrics in value as they yielded a higher number of awardees.

As useful as these studies are, they illustrate one major gap in the available literature. Most studies utilize fragmented datasets, rendering the setting of global standards for scientific output almost impossible. This proves that there is a need for more thorough research that analyses the broadest possible scope of parameters within a single-domain dataset. This research should be designed to rank these parameters systematically, determine which measures control the system, and establish new indices which more accurately attribute impact to researchers. By applying a less reductionist approach along with deepening the examination of these indices, researchers and politicians can deal with improving systems of assessing scientific contributions that are objective and just. More concentration in this area will ensure the setting of universal standards which balance quantity and quality of research output and consequently provide meaningful assessments of academic achievement.

Nonetheless, as Waltman (2016) argues, the design of bibliometric indicators should be grounded not only in technical or empirical strength, but in a clear theoretical understanding of what is being measured and why [10]. Similarly, Abramo and D'Angelo (2018) emphasize that the concept of "impact" must be rigorously defined before any metric can be meaningfully applied [11]. We fully agree with this viewpoint and acknowledge that citation-based metrics, while widely used, must be interpreted cautiously and within context. In response, our work integrates empirical ranking with external validation based on prestigious awards, providing an indirect operationalization of impact that complements, rather than replaces, deeper theoretical models. Furthermore, we recognize that the global conversation on research evaluation is moving toward more holistic, qualitative, and responsibility-driven models, as highlighted by initiatives like CoARA and critiques raised by Torres-Salinas et al. (2023) [12]. While our study remains within the quantitative domain, we view it as a stepping stone toward hybrid evaluation frameworks that combine empirical evidence with qualitative judgment, offering tools to inform-not dictate-research assessment policies.

3 Methodology

Building on insights from an extensive literature review, the focus now shifts to proposing a new index for evaluating researchers in the computer science field. Figure 1 presents the architecture diagram, illustrating the key stages of the methodology. The subsequent sections provide a detailed explanation of each stage, emphasizing their significance and contribution to the development of the proposed index.

3.1 Dataset collection

For the evaluation of proposed methodology, a domain-specific dataset was essential. For this study, we selected the field of computer science due to its rich research history, substantial contributions to innovation, and rapid evolution. Computer science plays a pivotal role in developing new technologies, methodologies, and solutions, making it a compelling subject for analysis. Its dynamic nature has also been the focus of prior research [33].

Our dataset consisted of 1,200 records, equally divided between awardees and non-awardees, with 600 entries in each category. Awardee data were sourced from prestigious organizations such as the Association for Computing Machinery (ACM) and IEEE, which recognize outstanding researchers through distinguished awards, including the ACM Turing Award, Computer Pioneer Award, IEEE John von Neumann Medal, and IEEE Technical

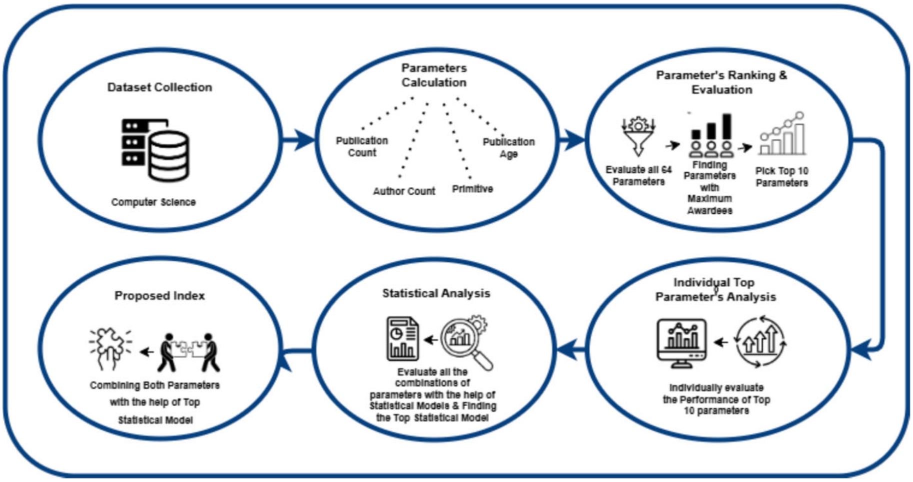

Fig. 1 The diagram illustrates the proposed framework, which is structured into multiple phases. Initially, the dataset is collected, followed by the identification and computation of existing bibliometric parameters. Authors are then ranked based on the calculated indices. In the next phase, the top 10 influential parameters are identified and all possible combinations of these parameters are generated. Finally, various statistical models are applied to combine the top parameter pairs, leading to the formulation of a new proposed index

Table 1 Dataset Statistics

Researchers metadata	Count
Authors count	1200
Awardees count	600
Non awardees count	600
Citation count	32,801,476
Publication count	171,388

Achievement Award. Non-awardee data were adapted from datasets previously used by Samreen et al. (2020) [33]. A detailed statistic is provided in Table 1.

Awardee information, including names and corresponding award years, was collected from society websites spanning the past three decades. Researcher data were extracted using the Publish or Perish platform, which employs advanced algorithms to retrieve metadata from Google Scholar. A ’hold on’ strategy was implemented to gather data prior to the award year, ensuring temporal consistency. To maintain dataset fairness and balance, non-awardees were selected in proportion to the number of awardees for each year. For example, if 15 awardees were recognized in 1999, data from 15 non-awardees before 1999 were also collected.

Prior to analysis, a thorough data preprocessing phase was carried out, especially for data collected from Google Scholar medium. This step was essential to eliminate errors and irrelevant content often termed as "noise" that could affect the reliability of the findings. The cleaning process included validating data accuracy and removing duplicate records. Furthermore, two critical procedures were applied to improve both the quality and relevance of the dataset: (i) a filtering mechanism ensured that research articles were strictly related to the computer science domain, and (ii) an author disambiguation process was applied, as previous studies have identified challenges in this area when working with Google Scholar data [9].

The dataset was divided into two categories: awardees & non-awardees. Moreover, names of the 600 awardees were extracted directly from society websites, eliminating the need for disambiguation. Non-awardee data were sourced from Samreen et al. (2020) [33]. Afterward, we applied the author disambiguation process, in which two key scenarios were identified: (i) cases where researchers had identical first and last names, requiring verification, and (ii) cases where different individuals shared the same last name but had distinct first names, necessitating further evaluation. These cases were resolved using established methodologies from the literature [34, 35]. No identical first and last names were found among the 600 non-awardees, making the first scenario inapplicable. However, 33 cases involved authors with the same last name, requiring further analysis. The evaluation confirmed that 37 out of 51 authors with matching last names but different first names were distinct individuals, while 20 out of 49 were variations of the same person. To maintain dataset balance, additional unique authors were incorporated.

Moreover, the Publish or Perish tool proved invaluable in the disambiguation process, as its advanced search capabilities facilitated the identification of name variations, improving the accuracy and comprehensiveness of the dataset.

3.2 Parameter identification and calculation

In this study, we compiled a comprehensive set of 64 author-level bibliometric parameters through an extensive review of the existing literature on research evaluation. These parameters were selected to capture diverse aspects of research performance and scholarly impact. For clarity and structured analysis, the parameters were systematically classified into four categories: (i) primitive parameters, (ii) parameters based on publication and citation counts, (iii) parameters related to author count, and (iv) parameters considering the age of publications. Using the available dataset of computer science researchers, we calculated the value of each parameter for every individual in the dataset. This categorization enables a multi-dimensional evaluation of author performance and supports domain-specific analysis of impact, productivity, and influence. A visual overview of the parameters and its categories is provided in Fig. 2.

3.3 Evaluating parameters with simple ranking techniques

To assess the effectiveness of each of the 64 bibliometric parameters, we apply a straightforward ranking approach. For every parameter, a separate ranked list of all 1200 authors is generated by sorting them in descending order based on their index values. We then examine how many award recipients appear in the top 100 positions of each list. This method provides a direct measure of an index's ability to highlight influential researchers.

By comparing these rankings, we identify which indices most accurately reflect recognized scholarly excellence. The top ten parameters that capture the highest number of awardees within the top 100 are selected for deeper analysis. This process helps isolate the most promising indicators and ensures that our evaluation framework focuses on indices with strong discriminative power in the domain of computer science.

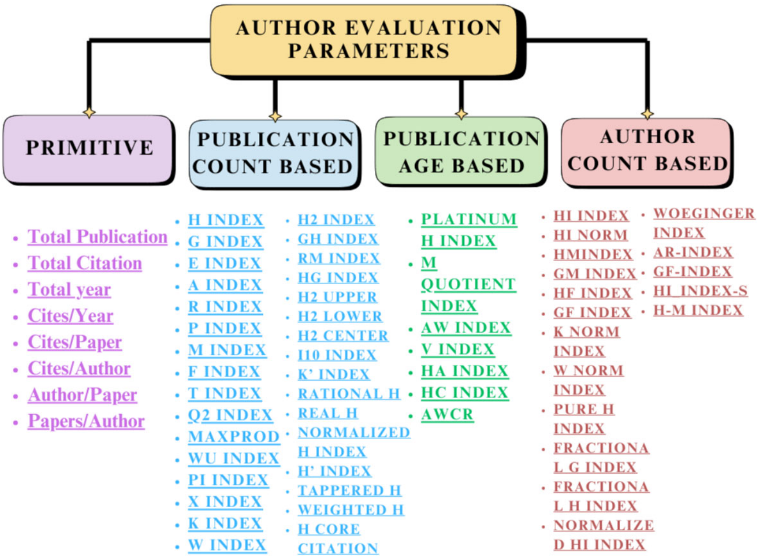

Fig. 2 Authors assessment parameters

3.4 Combined statistical analysis and proposed index

To effectively combine the top-performing individual parameters, we employed a range of statistical models aimed at capturing the collective strength of index pairs. Using the binomial coefficient, we generated 45 unique two-parameter combinations. Each pair was then evaluated using multiple statistical techniques, including the arithmetic mean, harmonic mean, geometric mean, contra-harmonic mean, quadratic mean, cubic mean, logarithmic mean, Lehmer mean, root-mean-square (RMS), and trigonometric mean. These models were selected for their mathematical diversity in aggregating values, allowing for a nuanced assessment of combined bibliometric indicators. Through this comparative analysis, we identified the most suitable statistical model for combining high-performing parameters. This model forms the foundation for the proposed index, with further validation discussed in the Results section. Detailed descriptions and formulations of each statistical method are provided in Table 2.

3.5 Proposed index validation

Validating the proposed index is essential to ensure its reliability and relevance in measuring scholarly impact. While traditional validation methods often rely on institutional datasets or simulated scenarios involving distinguished faculty, our study adopts a broader strategy by correlating the proposed index with recipients of prestigious international research awards. These awards, recognized across disciplines and institutions, serve as credible markers of excellence and peer acknowledgment. Although this validation method may not capture all dimensions-particularly contributions from emerging fields or researchers yet to receive formal recognition-it offers a globally representative and externally validated benchmark. Despite its limitations, this approach enhances the generalizability of the index and aligns it

Table 2 Statistical Models and Their Formulas

Statistical Model	Formula
Arithmetic mean (AM)	$AM = \frac{a_1+a_2+\cdots+a_n}{n}$
Harmonic mean (HM)	$HM = \frac{n}{\sum_{i=1}^{n}\frac{1}{a_i}}$
Geometric mean (GM)	$GM = \left(\prod_{i=1}^{n} a_i\right)^{\frac{1}{n}}$
Quadratic mean (QM)	$QM = \sqrt{\frac{a_1^2+a_2^2+\cdots+a_n^2}{n}}$
Cubic mean (CM)	$CM = \sqrt[3]{\frac{a_1^3+a_2^3+\cdots+a_n^3}{n}}$
Contra-harmonic mean (CHM)	$CHM = \frac{\sum_{i=1}^{n} a_i^2}{\sum_{i=1}^{n} a_i}$
Logarithmic mean (LM)	$LM(a,b) = \frac{b-a}{\ln b - \ln a}, \quad \text{for } a \neq b$

with widely accepted standards of academic excellence, ensuring that the measure reflects impactful contributions beyond institutional or regional boundaries.

4 Results and discussion

This section provides a comprehensive evaluation of our proposed approach, offering both numerical and visual interpretations to facilitate deeper understanding. We critically examine the performance of bibliometric parameters and the derived indices, especially in identifying top-performing researchers in the domain of computer science.

4.1 Parameter evaluation through basic ranking methods

To systematically assess the performance of 64 bibliometric parameters within the computer science domain, we ranked all parameters based on the number of top 100 award-winning researchers they successfully identified. Parameters were sorted in descending order, and a focused analysis was conducted on the top 10 indicators. This strategy ensured balanced evaluation while emphasizing the most impactful metrics.

Figure 3 illustrates the top-ranking parameters. The *h2-upper index* emerged as the most effective, correctly identifying 72 award recipients. It was followed by the *k-index* (64 award recipients) and *Cite per Paper* (63 award recipients). Other notable parameters included the *Pi-index* and *A-index* (each with 60 award recipients), and the *Ar index* (59 award recipients). Moderately effective indicators such as the *weighted h-index*, *h-core citations*, and *R-index* identified around 50 award recipients each. In contrast, weaker performers like the *k-bar index* and *h2-mean index* identified only 28 award recipients. Visual analytics aid in discerning this wide variation in metric effectiveness.

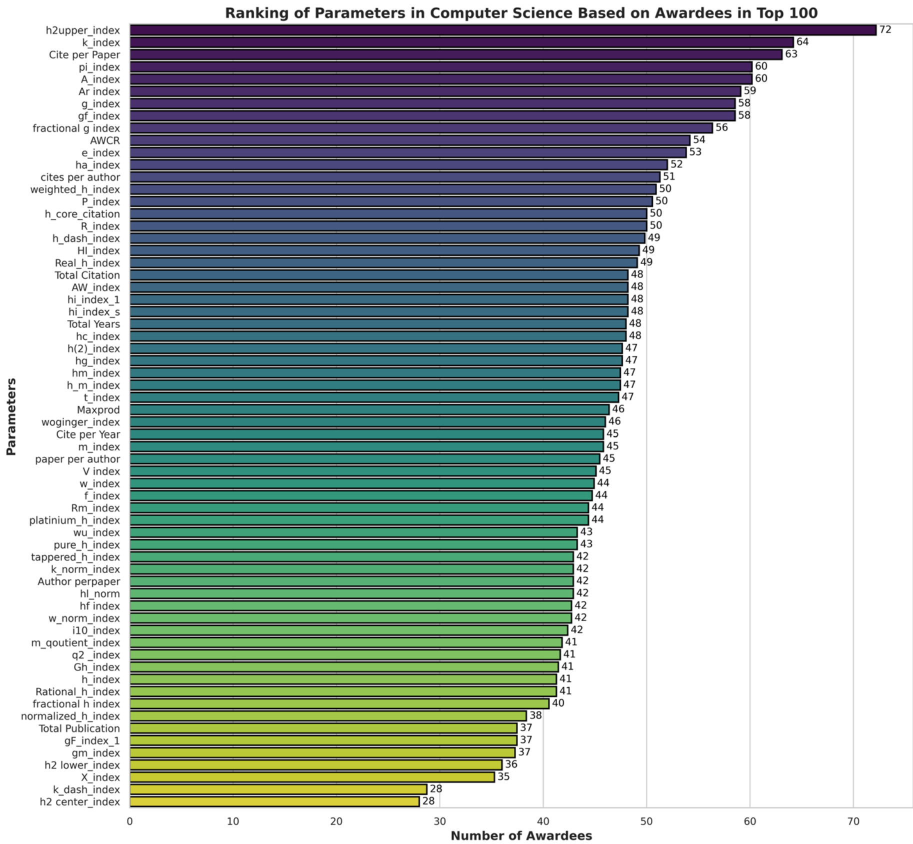

Fig. 3 Ranking of the parameters in computer science

4.2 Individual analysis of key parameters

In this section, we have assessed the 10 most significant parameters individually. Figure 4 presents a comparative visualization of the top 10 bibliometric parameters. The *h2-upper index*, *k-index*, and *Cite per Paper* achieved the highest weighted averages, confirming their strength in distinguishing influential researchers. Indicators such as the *Pi-index* and *A-index* reinforce the relevance of productivity-impact combinations, while *Ar index* and *g-index* capture broader citation patterns. Additionally, indices like the *gf-index*, *fractional g-index*, and *AWCR* reflect more nuanced, author-normalized impact measurements.

4.3 Integrating statistical analysis and the proposed index

To evaluate the effectiveness of various averaging approaches in measuring author impact within scientometrics, we conducted a comparative analysis using multiple statistical models applied to the *h2-upper* and *AWCR* indices. As shown in Fig. 5, the models examined include the arithmetic mean (AM), harmonic mean (HM), geometric mean (GM), quadratic mean (QM), cubic mean (CM), contra-harmonic mean (CHM), and logarithmic mean (LM). Among these, the CHM demonstrated the highest accuracy at 64.46%, while the HM yielded the

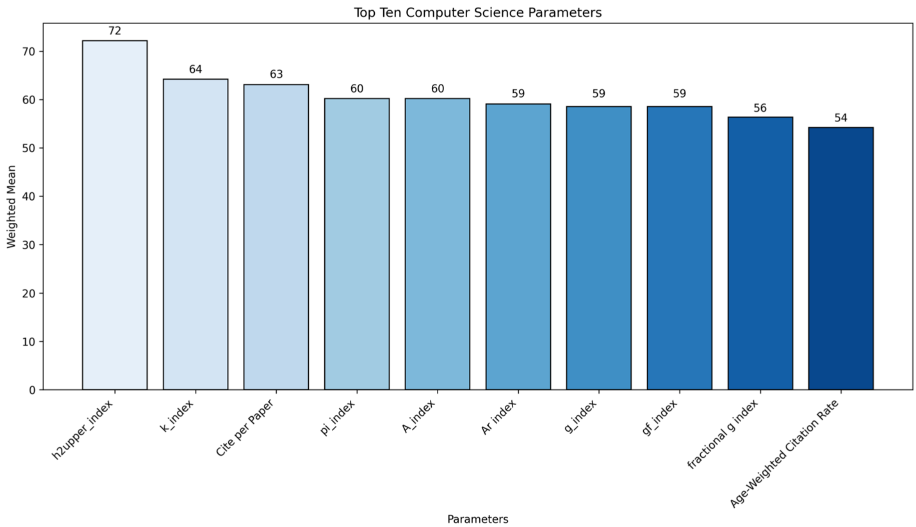

Fig. 4 Top 10 performing parameters based on the number of awardees ranked within the top 100 researchers. The figure highlights the top 10 parameters that most effectively identify high-impact individuals according to award recognition

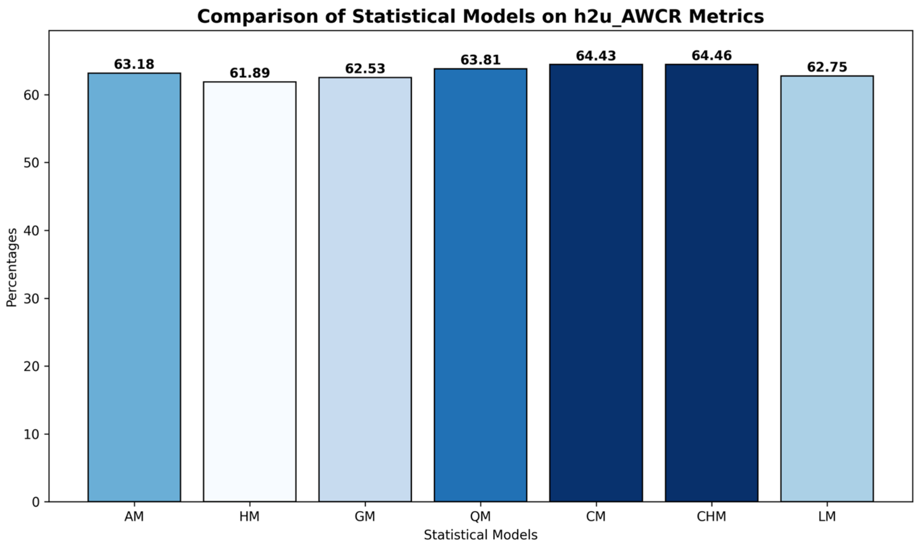

Fig. 5 Average impact analysis of the statistical models

lowest performance at 61.89%. The following subsections provide a detailed explanation of each model's underlying mathematical principles and discuss the outcomes of their respective calculations.

4.3.1 A.M

The arithmetic mean (AM) is a basic measure of central tendency, calculated by summing all values and dividing the total by the number of values to obtain an average. It is widely used in statistical analysis for its simplicity and ability to provide a general estimate of the overall trend in data. In the context of the *h2-upper* and *AWCR* indices, AM yields a value of 63.18% as shown in Fig. 5. While this result suggests that AM offers a reasonable estimate of the overall metrics, it falls short when accounting for the skewed distribution of citation data. Author impact data often exhibit significant variability, with a small number of highly cited papers contributing disproportionately to the overall impact. Because AM treats all values equally, it fails to capture the nuances of this distribution, making it less effective for accurately evaluating author impact in citation-based metrics.

4.3.2 H.M

The harmonic mean (HM) is specifically designed to give greater weight to smaller values in a dataset, making it particularly useful for evaluating ratios and rates. Unlike the arithmetic mean (AM), HM places more emphasis on lower values, ensuring that smaller contributions have a more significant impact on the final result. As shown in Fig. 5, HM yields the lowest performance at 61.89%, indicating that it is not an ideal choice for assessing the *h2-upper* and *AWCR* indices. This limitation arises because author impact data typically includes a mix of highly and less cited papers. In such cases, HM disproportionately diminishes the influence of highly cited works, ultimately failing to capture the true impact of highly productive researchers. As a result, the Harmonic Mean is not effective in providing an accurate reflection of author impact in citation-based metrics.

4.3.3 G.M

The geometric mean (GM) measures central tendency by considering the product of values rather than their sum, making it particularly useful for data that exhibits exponential growth or multiplicative effects, such as financial or citation-based metrics. As illustrated in Fig. 5, GM performs slightly better than the harmonic mean (HM), achieving 62.53% of the values. While GM effectively reduces the influence of extreme values more than the arithmetic mean (AM), the inherent nature of citation distribution still causes it to underperform when applied to the *h2-upper* and *AWCR* indices. Highly cited authors often make substantial contributions to their fields, and by treating all values equally, GM fails to adequately capture the influence of these highly impactful researchers. As a result, while GM provides a more balanced estimate than AM, it does not fully reflect the contribution of top-tier authors.

4.3.4 Q.M

The quadratic mean (QM), also known as the root-mean-square (RMS), assigns greater weight to higher values, making it particularly useful in scenarios where larger values play a more significant role in the overall score. This is especially relevant in contexts such as energy calculations or assessments of author impact, where the contributions of highly influential researchers should be more prominently recognized. As shown in Fig. 5, QM achieved a notable score of 63.81%, surpassing the performance of the arithmetic mean (AM), harmonic mean (HM), and geometric mean (GM). This outcome suggests that models which emphasize larger values tend to be more effective in scientometric applications. Given that citation distributions often feature extreme values particularly for highly cited authors the ability of the quadratic mean to amplify these values makes it a more appropriate choice than simpler arithmetic methods.

4.3.5 C.M

The cubic mean (CM) extends the concept of the quadratic mean (QM) by placing even more emphasis on higher values. This makes it particularly advantageous when the contributions of larger values such as those from highly cited authors should have a greater impact on the overall score. As depicted in Fig. 5, CM achieves an impressive score of 64.43%, marking a significant improvement over the previous models. The success of CM lies in its ability to effectively differentiate between highly cited researchers and those with lower citation counts. This is crucial for the h2-upper and AWCR index, which is designed to identify influential researchers based on their overall contributions. The ability of CM to highlight the most impactful researchers aligns with the goals of these indices, which aim to assess the true extent of an author's scholarly influence.

4.3.6 C.H.M

As shown in Fig. 5, the Contra Harmonic Mean (CHM) Index excels with a performance of 64.46%, outperforming all other models used for calculating the h2-upper and AWCR index. CHM's success lies in its ability to balance the influence of both small and large values while effectively identifying highly cited authors, even those with fewer citations. This flexibility is particularly valuable when dealing with complex, citation-based indices. The CHM's design, which incorporates squared values in both the numerator and denominator, allows it to account for contributions across the entire range of data. Its superior performance indicates that CHM's adaptability to variability and extreme values outperforms simpler models, such as the arithmetic mean (AM) and geometric mean (GM), both of which rely on mean values and fail to capture the nuances of author impact as effectively.

4.3.7 L.M

The logarithmic mean (LM) is commonly used in cases involving exponential distributions or logarithmic growth. It is particularly suited for situations where the focus is on proportionate rather than absolute differences between values. With a score of 62.75% in Fig. 5, LM demonstrates moderate effectiveness in capturing citation trends. While it performs better than the arithmetic mean (AM) model by accounting for shifts in citation distribution, it falls short compared to models like the contra-harmonic mean (CHM) and cubic mean (CM). This suggests that while logarithmic transformation captures some aspects of citation behavior, it does not fully capture the complexities required by the h2-upper and AWCR indices.

4.4 Performance of statistical models on h2-upper and AWCR index

The contra-harmonic mean (CHM) model, which integrates the h2-upper and age-weighted citation rate (AWCR) indices, outperforms all other models. This is due to its ability to emphasize higher contributions while capturing complex citation relationships. The results, as shown in Fig. 5, demonstrate that CHM achieves the highest impact at 64.46%, followed closely by the cubic mean (CM) at 64.43% and the quadratic mean (QM) at 63.81%. These models prioritize larger values, making them particularly suitable for evaluating researcher impact based on significant contributions. On the other hand, the harmonic mean (HM) and logarithmic mean (LM) show weaker performances at 61.89% and 62.75%, respectively, due to their reliance on smaller values, which distorts the overall assessment of influential researchers.

The arithmetic mean (AM), which treats all values equally, has a moderate impact of 63.18%. This suggests that models giving greater weight to larger contributions, such as CHM, CM, and QM, are more effective in scientometric applications, particularly for indices like h2-upper and AWCR.

4.5 Implications for research performance evaluation and model limitations

While the CHM model demonstrates superior performance, its direct integration of the AWCR and h2-upper indices assumes equal contribution from both, which oversimplifies their predictive power. This assumption may not fully reflect the true impact of these indices on research outcomes. Figure 5 illustrates the average impact of statistical models on the two parameters, emphasizing that an equal contribution approach may lead to a less accurate representation.

To address this limitation, we propose a refined weighting scheme that better reflects the relative significance of each index. This scheme assigns higher weights to more influential indices, allowing for a more balanced calculation in the final model. The weight for each index is calculated using the following normalization formula:

$$w_i = \frac{x_i}{\sum\limits_{i=1}^{n} x_i}$$

where x_i represents the value of each index. After calculating the weights, the AWCR and h2-upper indices are combined using the contra-harmonic mean to form a new, more comprehensive index, referred to as the hA-index. The hA-index integrates the strengths of both indices and provides a more accurate evaluation of researcher impact.

The formula for the hA-index is:

$$hA \text{ Index} = \frac{(h2\text{-upper}(57.12) + \text{awcr}(42.88))^2}{h2\text{-upper}(57.12) + \text{awcr}(42.88)}$$

4.6 Theoretical justification and novelty of the hA-index

The *hA-index* is proposed as a domain-optimized metric for evaluating researcher impact by combining the h2-upper index and the age-weighted citation rate (AWCR). These two parameters were selected based on a comprehensive ranking of 64 author-level metrics, where h2-upper and AWCR consistently demonstrated the highest ability to identify award-winning researchers in the field of computer science.

To determine the most suitable method for combining these metrics, we systematically evaluated seven statistical aggregation models, arithmetic mean (AM), harmonic mean (HM), geometric mean (GM), quadratic mean (QM), cubic mean (CM), contra-harmonic mean (CHM), and logarithmic mean (LM)—using all top-performing parameter pairs. Our analysis revealed that the CHM consistently outperformed the other models, achieving the highest accuracy (66.46%) in ranking award-winning researchers.

Importantly, CHM was not selected arbitrarily; its selection was data-driven, supported by empirical results. CHM emphasizes higher values in skewed distributions, a common feature in bibliometric data, and thus is well-suited to differentiate high-impact researchers from their peers. This allows the metric to address limitations of traditional indices, such as the h-index, which often undervalue a researcher's most influential works.

The final hA-index uses the CHM to integrate h2-upper and AWCR with a weighting scheme based on the relative ranking performance of each metric. The weighted formula is expressed as:

$$\text{hA-index} = \frac{(w_1 \cdot h2)^2 + (w_2 \cdot AWCR)^2}{w_1 \cdot h2 + w_2 \cdot AWCR} \tag{1}$$

where $w_1 = 57.12$ and $w_2 = 42.88$ reflect the normalized importance of h2-upper and AWCR in the ranking results. These weights were derived from empirical performance metrics that evaluated each index's ability to rank awardees within the top 100 researchers.

The novelty of the hA-index lies in its three-step construction process:

1. Data-driven selection of the strongest indicators (h2-upper and AWCR),
2. Empirical evaluation of multiple statistical models to determine the most effective mean (CHM),
3. Performance-based weighting to optimize the contribution of each component.

This methodology ensures that the hA-index is both theoretically grounded and empirically validated. It provides a nuanced, field-sensitive alternative to traditional indices by addressing citation skew, author contribution, and time-based influence—offering a more balanced and accurate reflection of researcher impact in computer science.

5 Conclusion

This study introduces a novel evaluation framework for researchers that enhances the precision and reliability of impact evaluation in the computer science domain. By systematically analyzing 64 bibliometric parameters and exploring their statistical combinations, we have identified the most effective indices for ranking researchers. Our findings demonstrate that the h2-upper index and the AWCR index exhibit the highest effectiveness, while the contra-harmonic mean outperforms other aggregation models in maintaining the distinctive properties of top-ranking parameters. The proposed index, derived using the contra-harmonic mean of the pair of best-performing parameters. By validating our approach against award-based benchmarks, we ensure that the proposed metric effectively differentiates high-impact researchers from their peers. This work contributes to the ongoing discourse on research evaluation by addressing the limitations of traditional metrics and offering a more comprehensive ranking methodology. The findings have significant implications for academic institutions, funding bodies, and policymakers, providing a refined approach for identifying influential researchers in computer science. Although the proposed index demonstrates substantial improvements in the evaluation of researchers, several avenues remain open for further exploration. First, future studies can extend this framework to other scientific domains to assess its generalizability. The relative impact of different indices can vary across disciplines, necessitating domain-specific adaptations. Second, integrating machine learning techniques into researcher ranking systems could further enhance predictive accuracy and robustness. Exploring deep learning models for automated feature selection and weight optimization presents a promising direction. We also acknowledge the philosophical concern that "*Not everything* *that counts can be counted*." While our analysis includes a large set of measurable parameters, our goal was not to assign equal importance to each but to empirically test their utility in identifying externally validated research excellence. The inclusion of 64 parameters enabled a comparative filtering process, where only the most meaningful and high-performing indicators were retained. Our intent is not to reduce scholarly impact to numbers, but to provide complementary evidence that may support more holistic, qualitative, and context-aware evaluation models. In addition, incorporating qualitative aspects such as peer reviews, editorial roles, and interdisciplinary collaborations into the ranking process could provide a more holistic evaluation of the influence of the researcher. Future work can also examine the temporal evolution of researcher rankings to assess long-term impact trajectories. Lastly, refining the validation process by incorporating more diverse award datasets and alternative benchmarking strategies can further strengthen the proposed methodology. By addressing these aspects, future research can continue improving the precision and fairness of researcher evaluation, ensuring a more comprehensive and equitable recognition of scholarly contributions.

Author Contributions H.Z.M wrote the paper, M.T.A and G.M supervise the study.

Data Availability This review has no associated data.

Declarations

Competing interests The authors declare no competing interests.

References

1. Bornmann L, Mutz R, Daniel H-D (2010) The h index research output measurement: two approaches to enhance its accuracy. J Inf 4(3):407–414
2. Bihari A, Tripathi S, Deepak A (2023) A review on h-index and its alternative indices. J Inf Sci 49(3):624–665
3. Jiang X, Sun X, Zhuge H (2013) Graph-based algorithms for ranking researchers: not all swans are white! Scientometrics 96:743–759
4. 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. Scient Rep 11(1):21900
5. Dorta-Gonzalez P, Dorta-González M-I (2013) Impact maturity times and citation time windows: the 2-year maximum journal impact factor. J Inf 7(3):593–602
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
7. Wilsdon J, Allen L, Belfiore E, Campbell P (2015) Johnson,: The metric tide: report of the independent review of the role of metrics in research assessment and management. HEFCE, Bristol
8. James, D (2014) Completing hirsch's h-index measuring scholarly impact. Scholardox E4
9. Harzing A-WK, Wal R (2008) Google scholar as a new source for citation analysis. Ethics Sci Environ Politics 8(1):61–73
10. Waltman L (2016) A review of the literature on citation impact indicators. J Inf 10(2):365–391
11. Abramo G, D'Angelo CA (2018) Reproducible and replicable bibliometric research: an opinion paper. J Inf 12(3):365–368
12. Torres-Salinas D, Moed HF, Rafols I (2023) The future of research assessment: How can we use metrics responsibly? Scientometrics 128(3):1215–1234
13. Bogers T, Bosch A (2008) Recommending scientific articles using citeulike. In: Proceedings of the 2008 ACM conference on recommender systems, pp. 287–290
14. Mahbuba D, Rousseau R (2013) Year-based h-type indicators. Scientometrics 96:785–797
15. Burrell QL (2007) On the h-index, the size of the hirsch core and jin's a-index. J Inf 1(2):170–177
16. Hirsch JE (2005) An index to quantify an individual's scientific research output. Proc Natl Acad Sci 102(46):16569–16572
17. Dienes KR (2015) Completing h. J Inf 9(2):385–397

18. Egghe L (2006) Theory and practise of the g-index. Scientometrics 69(1):131–152
19. Jin B, Liang L, Rousseau R, Egghe L (2007) The r-and ar-indices: complementing the h-index. Chinese Sci Bull 52(6):855–863
20. 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 Inf 4(1):23–28
21. Alonso S, Cabrerizo F, Herrera-Viedma E, Herrera F (2010) hg-index: a new index to characterize the scientific output of researchers based on the h-and g-indices. Scientometrics 82(2):391–400
22. Kosmulski M (2007) Maxprod-a new index for assessment of the scientific output of an individual, and a comparison with the h-index. Cybermetrics: Int J Scientometrics, Inf Bibliometrics (11), 5
23. 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
24. 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
25. Wan J-K, Hua P-H, Rousseau R (2007) The pure h-index: calculating an author'sh-index by taking co-authors into account. COLLNET J Scientometrics Inf Manag 1(2):1–5
26. Zhang C-T (2009) The e-index, complementing the h-index for excess citations. PloS One 4(5):5429
27. Katsaros D, Akritidis L, Bozanis P (2009) The f index: quantifying the impact of coterminal citations on scientists' ranking. J Am Soc Inf Sci Technol 60(5):1051–1056
28. Sidiropoulos A, Katsaros D, Manolopoulos Y (2007) Generalized hirsch h-index for disclosing latent facts in citation networks. Scientometrics 72(2):253–280
29. Wu Q (2010) The w-index: a measure to assess scientific impact by focusing on widely cited papers. J Am Soc Inf Sci Technol 61(3):609–614
30. Ameer M, Afzal MT (2019) Evaluation of h-index and its qualitative and quantitative variants in neuroscience. Scientometrics 121(2):653–673
31. 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
32. Ayaz S, Afzal MT (2016) Identification of conversion factor for completing-h index for the field of mathematics. Scientometrics 109(3):1511–1524
33. Ayaz S, Masood N (2020) Comparison of researchers' impact indices. PLoS One 15(5):0233765
34. 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
35. Ayaz S, Afzal MT (2016) Identification of conversion factor for completing-h index for the field of mathematics. Scientometrics 109(3):1511–1524

Hafiza Zarafshan Mukhtiar has completed her Master's degree in Data Science from Shifa Tameer-e-Millat University, Islamabad. She holds a Bachelor's degree in Statistics from Islamia College Peshawar and has developed a strong foundation in statistical analysis, machine learning, and R programming. Zarafshan has previously worked as a Data Analyst Intern at COMSATS University Islamabad and served as an Administrator at The Quaid Public School. Her research interests include statistical modeling, hypothesis testing, and the application of data science in public health and education. She is proficient in multiple languages including English, Urdu, and Pashto, and has participated in several workshops and conferences related to statistical science and digital skills development.

Ghulam Mustafa is an accomplished academic and researcher with a distinguished career in the field of Computer Science. He holds a PhD from the University of Engineering and Technology, Taxila, Pakistan, with a focus on data mining, information retrieval, and artificial neural networks. Dr. Mustafa earned his MS in Computer Science with distinction from the Capital University of Science and Technology, Islamabad, where he was awarded a Gold Medal for outstanding academic performance. He also received multiple Gold Medals during his Bachelor's degree in Software Engineering from COMSATS University, Abbottabad. With over seven years of experience, Dr. Mustafa currently serves as a Senior Lecturer at Shifa Tameer-e-Millat University, Islamabad. He has previously held roles as an Associate Lecturer and Junior Lecturer at Capital University of Science and Technology. Alongside his academic roles, Dr. Mustafa has industry experience as a Web Developer and Android Developer at A&F Solution Software House and has worked as a freelancer on various programming projects. His research contributions include over 50 publications with a cumulative impact factor of 53.2, focusing on topics such as author ranking indices, publication metrics, and machine learning applications in academic assessment. Dr. Mustafa is also involved in supervising PhD and MS students, contributing to the development of novel methodologies in scientific impact evaluation and academic research.

Muhammad Tanvir Afzal is an accomplished academic and researcher with extensive experience in the field of computer science. Currently, he serves as the Director of the Shifa School of Computing, a position he has held since February 2022. Dr. Afzal has a distinguished career, with over 20 years of experience in academia, having held various leadership roles in renowned institutions across Pakistan. These include serving as a Professor and Director of the Quality Enhancement Cell (QEC) at Namal Institute Mianwali, a Professor at the Capital University of Science and Technology, and an Associate Professor at Mohammad Ali Jinnah University. Dr. Afzal's academic journey began with his M.Sc. from Quaid-i-Azam University in Islamabad, where he earned a Gold Medal in 2004. He then pursued a PhD in Computer Science at Graz University of Technology in Austria, under the mentorship of the esteemed Professor Dr. Hermann Maurer, graduating with the highest distinction in 2010. Beyond his teaching, Dr. Afzal is an active contributor to the academic community. He holds several honorary positions, including serving as an Honorary Professor at the International Institute for Computational Mathematics (IICM) in Austria. He is the Editor-in-Chief of the Journal of Universal Computer Science, one of the oldest and most respected journals in the field, and has served on the boards of numerous academic institutions. Dr. Afzal's research interests are diverse and encompass data science, information retrieval, information visualization, scientometrics, and human factors engineering. He has authored over 120 journal publications, with a total impact factor exceeding 116. His work has been instrumental in advancing the understanding of scientific knowledge extraction, citation analysis, and research paper ranking systems. In addition to his academic achievements, Dr. Afzal is a frequent contributor to national and international conferences, where he has published numerous papers and book chapters. He has also been a key member of various boards and committees in Pakistani universities, including selection boards, research committees, and curriculum review committees. Dr. Afzal's expertise and leadership extend to mentorship, having supervised numerous graduate students and acted as an examiner for PhD and Master's theses at top institutions. His deep commitment to academic excellence and his contributions to research continue to inspire and shape the field of computer science in Pakistan and beyond.