# Toward the consolidation of a multi-metric-based journal ranking and categorization system for computer science subject areas

**Abdul Hameed; Muhammad Omar; Muhammad Bilal; Han Woo Park**

Recommended citation:

**Hameed, Abdul**; **Omar, Muhammad; Bilal, Muhammad**; **Park, Han Woo** (2023). "Toward the consolidation of a multi-metric-based journal ranking and categorization system for computer science subject areas". *Profesional de la información*, v. 32, n. 7, e320703.

*https://doi.org/10.3145/epi.2023.dic.03*

Manuscript received on 19th May 2023
Accepted on 29th June 2023

**Abdul Hameed**
*https://orcid.org/0000-0003-3235-7779*

*The Islamia University*
*Department of Computer Science*
Bahawalpur, Pakistan
*abdulhameedattari@gmail.com*

**Muhammad Omar**
*https://orcid.org/0000-0002-7071-5760*

*The Islamia University*
*Department of Computer Science*
Bahawalpur, Pakistan
*m.omar.nazeer@gmail.com*
(co-first author, and co-corresponding author)

**Muhammad Bilal**
*https://orcid.org/0009-0004-4757-3153*

*The Islamia University*
*Department of Computer Science*
Bahawalpur, Pakistan
*m_bilalcsiub@yahoo.com*

**Han Woo Park**
*https://orcid.org/0000-0002-1378-2473*

*YeungNam University*
No. 257 Humanities Hall
280, Dae-dong, Gyeongsan-si
38541 Gyeongsangbuk-do, South Korea
*hanpark@ynu.ac.kr*
(corresponding author)

**Abstract**

The evaluation of scientific journals poses challenges owing to the existence of various impact measures. This is because journal ranking is a multidimensional construct that may not be assessed effectively using a single metric such as an impact factor. A few studies have proposed an ensemble of metrics to prevent the bias induced by an individual metric. In this study, a multi-metric journal ranking method based on the standardized average index (SA index) was adopted to develop an extended standardized average index (ESA index). The ESA index utilizes six metrics: the CiteScore, Source Normalized Impact per Paper (SNIP), SCImago Journal Rank (SJR), Hirsh index (H-index), Eigenfactor Score, and Journal Impact Factor from three well-known databases (*Scopus*, *SCImago Journal & Country Rank*, and *Web of Science*). Experiments were conducted in two computer science subject areas: (1) artificial intelligence and (2) computer vision and pattern recognition. Comparing the results of the multi-metric-based journal ranking system with the SA index, it was demonstrated that the multi-metric ESA index exhibited high correlation with all other indicators and significantly outperformed the SA index. To further evaluate the performance of the model and determine the aggregate impact of bibliometric indices with the ESA index, we employed unsupervised machine learning techniques such as clustering coupled with principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). These techniques were utilized to measure the clustering impact of various bibliometric indicators on both the complete set of bibliometric features and the reduced set of features. Furthermore, the results of the ESA index were compared with those of other ranking systems, including the internationally recognized *Scopus*, SJR, and *HEC Journal Recognition System* (*HJRS*) used in Pakistan. These comparisons demonstrated that the multi-metric-based ESA index can serve as a valuable reference for publishers, journal editors, researchers, policymakers, librarians, and practitioners in journal selection, decision making, and professional assessment.

**Keywords**

Journal rankings; Research evaluation; Indicators; Scientific journals; Metrics; Algorithms; Machine learning; Cluster analysis; Principal component analysis (PCA); t-distributed stochastic neighbor embedding (t-SNE); Cross tabulation; ESA index.

**Competing interests**

The authors declare no competing interests.

**Acknowledgements**:

This work is related to the first author's PhD thesis.

## 1. Introduction

Evaluating research quality is a complex task that can significantly impact multiple decisions, such as improving the tenured track or basic pay scale systems (TTS/BPS) service structure to enhance research quality, determining hiring decisions, allocating research funding, conducting promotions, and awarding scholarly degrees.

Various evaluation systems have been developed for this purpose. Research standards can be evaluated through qualitative methods (**Wical**; **Kocken**, 2017), quantitative methods (**Beliakov**; **James**, 2011), or a hybrid approach that combines both methods (**Hsu** *et al.*, 2015). Additionally, a meta-approach for predicting journal quality has also been proposed (**Saarela**; **Kärkkäinen**, 2020). Conventionally, most journals assess the quality of a publication through a peer review process by experts in the relevant field of research (**Morris** *et al.*, 2009).

In the present era of information technology, various contemporary systems for ranking journals have been adopted by different organizations. *Web of Science* (*WoS*), *SCImago Journal & Country Rank*, and *Scopus* are the few examples of the numerous groups and for-profit institutions that maintain sizable publishing datasets that allow for the computation of citations and other potential journal influence statistics. Several well-defined bibliometric indicators have been developed for ranking journals, such as the Impact Factor (IF), Eigenfactor (EF) Score, Hirsh index (H-index), SCImago Journal Rank (SJR), Source Normalized Impact per Paper (SNIP), and CiteScore. Each metric has its own strengths and weaknesses. The IF is one of the most widely used indicators for ranking journals. However, the use of an individual indicator does not ensure reliable results (**Setti**, 2013). The main problem with citation-based indicators such as the IF, is the dissimilarity in citation practices among different disciplines. For instance, mathematical studies tend to receive fewer citations than biology research (**Ferrer-Sapena** *et al.*, 2016). To address this issue, bibliometric analysis has been applied to assess the influence of published work and their potential to enhance a journal's reputation (**Perera**; **Wijewickrema**, 2018).

Recently, there has been growing interest in the use of machine learning algorithms to automatically categorize journals, although this approach is not yet widely adopted (**Abbas** *et al.*, 2019). These models are capable of operating independently without the need for human intervention, which offers the potential for objectivity. In addition, automated classification procedures are typically less expensive to implement as compared to expert-based classification procedures. Moreover, machine learning algorithms have the advantage of being able to consider all available quality indicators, unlike citation-based indicators which are limited in scope. This implies the utilization of all the feasible bibliometric indicators (**Perera**; **Wijewickrema**, 2018). This is necessary because journal ranking is a multicriteria decision problem.

Based on the journal impact index, SJR ranks journals in each subject category into quartiles ranging from one to four (Q1-Q4). Q1 represents the top 25% SJR distribution, Q2 denotes the middle-high SJR distribution (25%-50%), Q3 indicates the middle-low SJR distribution (50%-75%), and Q4 refers to the lowest SJR distribution (the bottom 25%) (**Mañana-Rodríguez**, 2015).

The *Higher Education Commission* (*HEC*) of Pakistan typically categorizes journals into four groups to ensure research quality: W, X, Y and Z, where W indicates the highest standard, and Z indicates the lowest. The *HEC Journal Recognition System* (*HJRS*) is a recently developed online system for recognizing journals. On introducing HJRS, the *HEC* removed the Z category. The categorization system designed by the *HEC* of Pakistan only recognizes research journals that fall into the W, X, and Y categories with full implementation starting in July 2020. The *HEC* asserts that this system assesses the quality of publications using internationally acclaimed parameters (**Mubarak**; **Seemee**, 2021).

This study proposes a data-driven methodology for automatically categorizing computer science journals based on various features (metrics). The research questions for this study are as follows:

(1) Can we adopt a multi-metric-based scientific journal ranking system to develop an index known as the extended standardized average index (ESA index) to combine a number of bibliometric indices that can yield more robust and aggregated journal rankings?

(2) What is the impact of multiple bibliometric features on journal ranking? To what extent does the ESA index correlate with other bibliometric indicators?

(3) To implement a cluster analysis of the considered bibliometric indices (in conjunction with the ESA index (seven indices)) against a reduced set of indices to assess the stability of the corresponding journal ranking and categorization system.

(4) Can the ESA index function as an authentic and reliable medium for classifying the quality tiers of the *Scopus* Quartiles, *SJR* Best Quartiles, and *HJRS* Categories?

The remainder of this article is organized as follows. In Section 2, the related work and background informations regarding various bibliometric indicators, their definitions, and their advantages and disadvantages are discussed. Section 3 explains the materials and methods. Section 4 presents the experimental observations obtained from each method. Section 5 summarizes the concluding remarks and presents the conclusions.

## 2. Related work

The rankings of academic publications have an impact on various players in academia. Scientists consider potential venues for their research based on the rankings, departments assess their productivity using these rankings, and funding success may also be influenced by them (**Wical**; **Kocken**, 2017). The impact of academic publication rankings is not limited to academia alone, as it also affects the non-academic world. This includes publishers who want to assess the reputation of their journals, professional bodies, practitioners, and funding agencies. The application of scientometric methods in science and technology studies (STS) (**Wyatt**; **Milojevič**; **Park**; **Leydesdorff**, 2017) has significant implications for research quality. Numerous countries have implemented journal assessment standards to encourage and incentivize national academic institutions and research centers to actively contribute to the knowledge base in their respective fields (**Holmberg**; **Park**, 2018; **Saarela** *et al.*, 2016). While some rely on qualitative assessments through peer review (**Wical**; **Kocken**, 2017), others use quantitative metrics (**Yuen**, 2018), and few utilize hybrid (**Allen** *et al.*, 2009) or meta-ranking approaches (**Ennas** *et al.*, 2015). When assessing scholarly output, the quantity and quality of publications should be considered (**Zhu**; **Park**, 2022). Assessment techniques are established to gather evidence and information that can be used to evaluate different aspects of research and make informed decisions.

It is difficult to design and evaluate a system that aims to translate research materials into monetary rewards. One could contend that if an evaluation criterion based on quantitative measurements is relatively straightforward, it can have negative consequences. Since its introduction, the IF has been commonly utilized as a quantitative research method. However, there are many restrictions related to its misapplication (**Dellavalle** *et al.*, 2007). Therefore, other indices such as the SJR (**González-Pereira** *et al.*, 2010), H-index (**Lacasse** *et al.*, 2011), Eigenfactor (**Bergstrom**, 2007), CiteScore, (**James** *et al.*, 2018), and SNIP (**Moed**, 2010) have become popular for research evaluation.

Another indicator that measures the article effect is the *Altmetric*. It is based on Internet attention (**Holmberg**; **Park**, 2018). The *Altmetric* score is a metric that measures online attentions received by scholarly articles based on mentions in news publications, blog comments, tweets, and social media posts. The *Altmetric* score is often used to identify publications that have attracted a lot of attentions on the Internet (**Holmberg**; **Park**, 2018).

The literature on journal quality evaluation can be classified into four categories: conventional *subjective* ranking (qualitative approach), which is based on the opinions of experts in a specific discipline; *objective* ranking (quantitative approach), which is based on citations; *hybrid* ranking (hybrid approach), which is a combination of subjective and objective rankings; and *meta* ranking approach, which automatically ranks journals using artificial intelligence.

### 2.1. Qualitative approach for journal ranking

A qualitative or survey-based approach involves ranking journals based on their perceived quality and reputation by receiving feedback from qualified experts to rank journals in a specific domain (**Allen** *et al.*, 2009; **Walters**, 2017). There are two main drawbacks. First, this measure is effective only at the time it is used. This is because in a dynamic research field the top-ranking journals and popular subjects change over time (**Duan** *et al.*, 2018). Second, the ranking lists produced by survey-based methods become increasingly less trustworthy for lower-ranking journals.

### 2.2. Quantitative approach for journal ranking

Using quantitative approaches, journals are assessed according to their size (number of publications), influence, and number of citations (**Leydesdorff**; **Park**, 2017). These techniques are utilized to evaluate the journal quality, although these capture only a few features of quality and are simple to compute. However, it should be noted that quantitative factors are occasionally unrelated to the qualitative factors. For instance, the mere fact that a paper is published in a journal with a high volume of publications does not guarantee its quality (**Fersht**, 2009; **Tsai**, 2014). The main features of the quantitative metrics used in this study are summarized in Table 1.

### 2.3. Hybrid methods for journal ranking

Professionals help in decision-making to overcome the inherent drawbacks of using an individual index while maintaining the advantages of utilizing various indices and providing a distinctive aggregate score. Quantitative approaches are straightforward, unbiased, and current methods. However, survey-based approaches can incorporate qualitative data that are difficult to measure, and provide a tiered structure that aids in the creation of guidelines. An increasing number of journal rating experts consider that combining journal bibliometrics with professional assessment of journal quality is the best overall approach (**Tüselmann** *et al.*, 2015).

Business schools frequently use the A*ssociation of Business Schools* (*ABS*) *Academic Journal Guide* from among several journal ranking lists produced using hybrid techniques (**Morris** *et al.*, 2009). To develop this journal guide, members of the *ABS Scientific Council* have provided various measures such as the IF, SNIP, and SJR for each journal. After con-

Table 1. Summary of the main features of journal impact indicators provided in *WoS*, *SCImago Journal & Country Rank*, and *Scopus*

| Characteristic | *WoS* | | *SCImago Journal & Country Rank* | | *Scopus* | |
|---|---|---|---|---|---|---|
| | JIF* | EF* | H-index* | SJR* | CS* | SNIP* |
| Calculation methodology | Ratio of citations and publications | Based on Eigenvector centrality | Ratio of citations and publications (h citations from h papers) | Citations network-based | Ratio of citations and publications | The ratio of publications to citations, normalized by citation densities across disciplines |
| Publication window (years) | 2/5 | 5 | h | 3 | 4 | 3 |
| Citation window (years) | 1 | 1 | 1 | 1 | 4 | 1 |
| Journal self-citations | Yes | No | Yes | limited up to 33% | Yes | Yes |
| Normalized by papers count in the journal (size independent) | Yes | No | Yes | Yes | Yes | Yes |
| Normalized by fields/disciplines | No | No | No | Not directly, Thematic closeness based between journals | No | Yes |
| Normalized by reputation (weighted) | No | Yes | No | Yes | No | No |
| Applicability | Only for *JCR* journals | Only for *JCR* journals | for journals in Google Scholar | for all sources (including journals, conference proceedings, book series and trade publications) | for all sources (including journals, conference proceedings, book series and trade publications) | for all sources (including journals, conference proceedings, book series and trade publications) |
| Availability | Requires a subscription to *JCR* | Requires a subscription to *JCR* | Free, (no subscription required) | Free, (no subscription required) | Free, (no subscription required) | Free, (no subscription required) |
| Limitations and drawbacks | Different types of documents included in numerator and denominator, potentially manipulable, Short citation window (for 2-Year JIF); not normalized for fields/disciplines | Inconvenient numerical value, decreasing with new journals as included in the database; not normalized for fields/disciplines | Differing citation practices of articles in different fields | Complex calculation, difficult to interpret | Not normalized for disciplines | Impact per paper but indicates impact of average articles in a journal (not for a specific article) |

*JIF, Journal Impact Factor; EF, Eigenfactor; H-index, Hirsh index; SJR, SCImago Journal Rank; CS, CiteScore; SNIP, Source Normalized Impact per Publication.

sultation with their individual academic communities, they were instructed to group each publication into one of the following five categories: 4* for elite journals, 4 for top journals, 3 for highly regarded journals, 2 for good standard journals, and 1 for modest journals.

## 2.4. Meta-ranking approach for journal ranking

Recently, the concept of automatically ranking journals using machine-learning techniques has attracted significant attention (**Halim**; **Khan**, 2019). Because machine-learning based algorithms can operate without human intervention, these appear to be more objective. The *HJRS* is based on the meta-ranking approach and considers all the available quality indicators, unlike citation-based indicators that only consider a limited set of explanatory features. This is advantageous because ranking academic journals involves multiple criteria and decision-making factors. In various studies, machine-learning techniques such as regularized logistic regression, gradient boosting, and random forest have been used to predict journal quality (**Saarela**; **Kärkkäinen**, 2020). As shown in Table 2, several studies have demonstrated that machine-learning techniques provide better results than the qualitative, quantitative, and hybrid approaches adopted earlier. The principal component analysis (PCA) by (**Bollen** *et al.*, 2009) indicates the multidimensionality of the different impact indicators, (**Ennas** *et al.*, 2015) used various statistical and machine learning techniques to formalize an approach that ranks journals from different dimensions, thereby characterizing the aspects of research quality. The ensemble simple linear regression model by (**Duan** *et al.*, 2018) performed better for the interdisciplinary journals. A few studies on journal rankings are summarized in Table 2.

Table 2. Summary of journal ranking studies using machine learning techniques

| Work | Study Purpose | Variables | Techniques | Findings |
|---|---|---|---|---|
| (**Bollen** *et al.*, 2009) | Evaluating research impact through citations and usage data sets. | 39 bibliometric indicators | PCA | The Principal components show 92% of the variances between the correlations of journal rankings by 37 impact measures |
| (**Tüselmann** *et al.*, 2015) | Handling missing values and journals by DEA | Impact Factor, ABS, ABDC, VBH, CNRS | Random Forest, DEA | Treatment of missing data through imputation and better classification of journals through Random Forest |
| (**Tsai**, 2014) | Ranking computer science journals using IF and H-index | IF, 5-IF, H-index, | CombSUM | Find a better correlation between the impact factor and H-index of computer science journals |
| (**Ennas** *et al.*, 2015) | A data-driven methodology using different methods from statistics and machine learning to combine various indices to create an aggregate rating. | IF, 5-IF, SJR, H-index, Immediacy Index, Eigenfactor Score. Article Influence, SNIP, IPP | SVR, CombSUM, Borda Count and PCA | SVR and PCA outperformed well in ranking journals |
| (**Fernández-Cano**; **Fernández-Guerrero**, 2017) | EM journals were subjected to a multivariate evaluation based on seven highly linked evaluation variables to produce a factor-based meta-index | IF, H-index, SJR and two altmetric scores (3 months and any time) | Cronbach's alpha | The length of time (number of years) that each journal has been published would be a significant factor regarding the H-index that should be taken into consideration. |
| (**Duan** *et al.*, 2018) | A data-driven method used to rank MIS journals not included in ABS list | 2-year IF, 5-year IF, EF, AI, SNIP, SJR, ABS, ABDC, VHB, CNRS, FNEGE | MLR, ESLR, SVM, NN | ESLR achieves the best performance among various data-driven methods and generates reasonable ranking for new journals, top journals and interdisciplinary journals |
| (**Perera**; **Wijewickrema**, 2018) | Investigates the relationship among four journal rankings | IF, Eigenfactor, H-index and SJR | Pearson correlation coefficient, Hierarchical clustering, PCA, Kaiser-Meyer-Olkin (KMO) test and Bartlett's test | Results indicate that a higher correlation was found between IF and SJR |
| (**Halim**; **Khan**, 2019) | Data Science framework to automatically categorize journals | 19 features (IF, CiteScore, SNIP, H-index, SJR, Eigenfactor, article influence, immediacy index, cited half-life, publisher, website, country, age, open access, citations, percentile, peer review, number of articles published yearly) and acceptance rate | Feature selection (MI, mRMR, SD)<br>Clustering (k-means, k-medoids),<br>Classification (ANN, KNN)<br>Cluster validation using DBI, DI, SC, CHI | Top nine features (CiteScore, H-index, SJR, SNIP, cited half-life, Eigenfactor, article influence, total citations, percentile) four clusters identified,<br>Average accuracy (ANN) 92.86% |
| (**Saarela**; **Kärkkäinen**, 2020) | Automated rankings based on the analysis of bibliometric indicators through the expert score ranking and through data analysis and machine learning techniques | Features used (Rank, Title, publications, volume, type, start year, Norway Score, Denmark Score, SJR, IPP, SNIP, Panel, Sherpa/ Romeo Code, Publisher | SMOTE, Logistic regression, random forest and gradient boosting | High correlation found between citations- and expert-based rankings system. |
| (**Feng** *et al.*, 2020) | Identified the most important and contributing features for categorizing journals through unsupervised Laplacian score | 2-Year IF, 5-Year IF, CiteScore, SNIP, SJR and H-index with two class labels (discipline and quantile) | Laplacian score, spectral clustering, *k*-NN, BPNN and subjective method (questionnaire used) | Based on experimental results IF, CiteScore, and H-index are the best features and by the voting method based on a seven-point Likert scale, Impact Factor and H-index got higher votes. |

PCA: Principal Component Analysis, DAE: Data envelopment analysis, ABDC: Australian Business Deans Council, ABS: Association of Business Schools, VHB: Association of University of Business in German-Speaking Countries, CNRS: Centre National de la Recherche Scientifique, IF: Impact Factor, SNIP: Source Normalized Impact per Publication, SJR: SCImago Journal Rank, SVM: Support Vector Machine, SVR: Support Vector Regression, NN: Neural Network, BPNN: Back Propagation Neural Network, MLR: Multicollinearity problem, ESLR: Ensemble Simple Linear Regression

## 3. Methodology

The proposed methodology is in line with the intellectual recommendations of Leydesdorff's research group, namely, the use of scientometric methods in science and technology studies (STS) (**Wyatt**; **Milojevič**; **Park**; **Leydesdorff**, 2017). This study proposed a data-driven methodology to develop a novel multi-metric-based scientific impact measure called the ESA index for ranking and categorizing journals based on various bibliometric impact measures. The main objective was to propose an automated approach for categorizing journals using machine learning techniques in various computer science disciplines. Various bibliometric indicators used for this purpose were as the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor. Each bibliometric measure has its advantages and drawbacks, and the rankings it produces can vary significantly depending on the specific metric used and the criteria for ranking. The integration of current bibliometric indicators is a potential strategy to compensate the limitations of individual indicators. A multi-dimensional space constructed using different impact measures was used to assess the journals. First, a multi-metric based ESA index was proposed for ranking and categorizing academic journals. The proposed ESA index was developed from multiple impact measures, which combines the impact of each bibliometric measure. Thereby, it functions as an alternative to various indicators for academic journal quality assessment. The journals in various disciplines of computer science were analyzed and categorized using various well-known bibliometric features (the CiteScore, SNIP, SJR, H-Index, Eigenfactor Score, and Journal IF). Consequently, we formulated a data-driven methodology to determine the impact of the ESA index with other bibliometric indicators using machine learning techniques. For this purpose, we first applied k-means clustering to the full featured dataset (seven bibliometric features). We then applied two dimensionality reduction techniques (PCA and t-distributed stochastic neighbor embedding (t-SNE)) to determine the reduced set of features. Subsequently, we applied k-means clustering to a reduced set of features. The clustering results of the proposed model were compared and validated using the two most commonly used and currently available benchmarks: (1) *SJR* Best Quartiles and (2) *Scopus* Quartiles. The proposed methodology for the preliminary investigation of journal evaluations is presented in Figure 1.

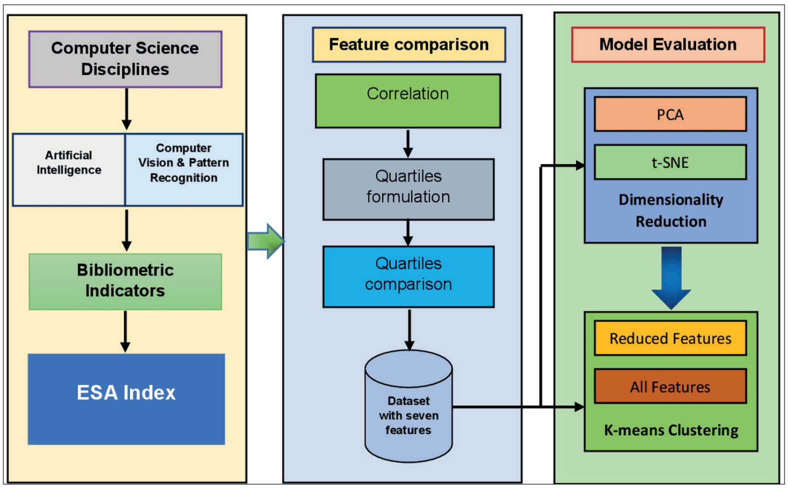


Figure 1. Block diagram of the proposed research methodology

The proposed framework utilizes unsupervised machine-learning approaches such as clustering and dimensionality reduction for journal evaluation. The following section discusses the various modules of the proposed system.

### 3.1. Dataset collection

The dataset consists of all the available journals indexed in *Scopus*, *Web of Science* (*WoS*), *SJR*, or *HJRS*. The dataset was extracted from various computer science disciplines to evaluate journal quality in their respective fields. A dataset currently available for 2021 was used in this study. The extracted features included are ISSN, Journal Title, CiteScore, SNIP, SJR, H-Index, Eigenfactor Score, and Journal IF. The *Scopus* Quartiles, *SJR* Quartiles, and *HJRS* journal categories were utilized to evaluate the validity of the proposed model. For a comprehensive analysis, datasets from two disciplines of computer science (314 journals of artificial intelligence, and 106 journals of computer vision and pattern recognition) were extracted with various bibliometric features. Various journal categories such as journal quartiles Q1-Q4 were extracted from the *SJR* and *Scopus* databases. Three journal categories (W, X, and Y) were extracted from the *HJRS*. A merging technique using outer join was applied on the collected dataset to ensure that journals indexed in any of the well-known databases such as *Scopus*, *Web of Science* (*WoS*), and *SJR* were included in the dataset. The distributions of journals in the *Scopus*, *SJR*, and *HJRS* categories are shown in Figures 2, 3, and 4, respectively.

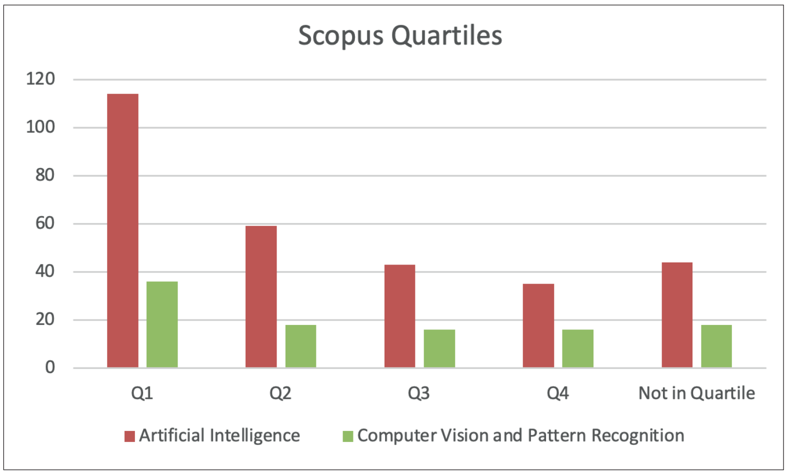


Figure 2. Distribution of journals with *Scopus* Quartiles

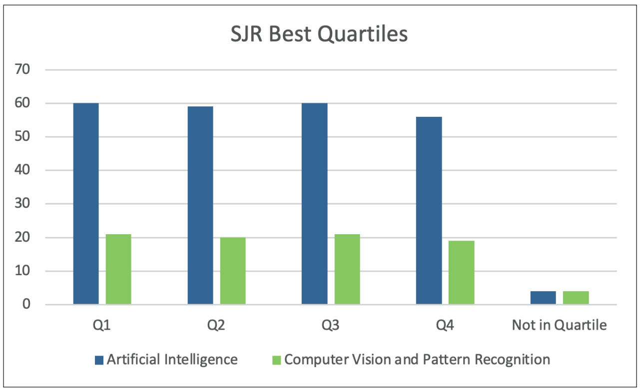



Figure 3. Distribution of journals with SJR Best Quartiles

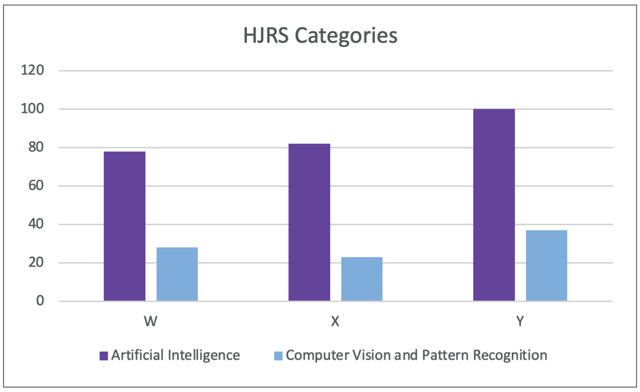



Figure 4. Distribution of Journals with HJRS categories

### 3.2. Preprocessing

After finalizing the dataset, standardization was implemented to preprocess the dataset for experimental use. The missing values of certain bibliometric indicators were filled with zero, and the dataset was standardized using a standard scalar. The standard scalar shifted the data of all the features in the range 0-1.

After removing the values with missing quartiles, 217 journals of artificial intelligence and 74 journals of computer vision and pattern recognition were used for data analysis, journal ranking, and categorization.

### 3.3. ESA index: Multi-metric based meta-ranking approach for journal ranking

Different bibliometric measures generally provide different journal rankings. This can cause ambiguities in the decision-making process. Therefore, an ESA index was introduced in this study using a feature engineering technique. The index is multi-metric because it combines various bibliometric features to propose a simple and reliable metric for ranking academic journals.

This study presented an approach to develop an ESA index as an alternative multi-metric impact indicator for evaluating academic journals. The aim was to contribute to multiple scientific impact measures such as the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor. These bibliometric indicators were combined to develop a new metric that is simple and multi-metric-based for journal evaluation. Various impact indicators have their advantages and shortcomings. Therefore, it is necessary to use multiple indicators rather than an individual one to evaluate the journal quality. However, owing to different calculation criteria, various impact metrics generally yield different evaluation results. Journal articles with a high IF do not necessarily have a high CiteScore and vice versa. Therefore, researchers would select only one of these journals as a reference for article submission. To develop an alternative, simple, and reliable metric for various impact indicators, we adopted a concept from the method used in (**Hsu** *et al.*, 2015).

The ESA index can be calculated as:

- Normalization: Compute the normalized value/score of each journal's research impact metric from the total score. The metrics used are the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor(IF).
- Average percentage: Calculate the average of the input features and determine the percentage.

For a given set of journals $D = \{x_i\}_i^n$ (where $x_i = (x_i^1, \ldots x_i^d) \in \mathbb{R}^d$ represents $d$ index values for the ith journal, e.g. $x_i^1, x_i^2, x_i^3, x_i^4, x_i^5$ and $x_i^6$ may, respectively represent the CiteScore, SNIP, SJR, H-index, IF, and Eigenfactor Score for a journal in the set $D$), the ESA index is calculated as follows:

(1) First, calculate the normalized value of each indicator as

(i) Calculate the normalized value of CiteScore as

$$nCS = \frac{CSi}{\sum_{i=1}^{n} CSi} \quad (1)$$

(ii) Calculate the normalized value of the SNIP as

$$nSNIP = \frac{SNIPi}{\sum_{i=1}^{n} SNIPi} \quad (2)$$

(iii) Calculate the normalized value of SJR as

$$nSJR = \frac{SJRi}{\sum_{i=1}^{n} SJRi} \quad (3)$$

(iv) Calculate the normalized value of the H-index as

$$nHI = \frac{HIi}{\sum_{i=1}^{n} HIi} \quad (4)$$

(v) Calculate the normalized value of the IF as

$$nIF = \frac{IFi}{\sum_{i=1}^{n} IFi} \quad (5)$$

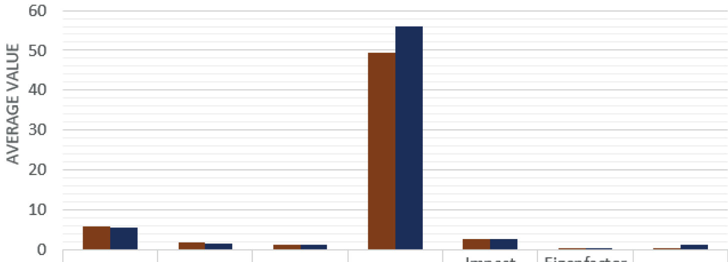
(vi) Calculate the normalized value of the Eigenfactor Score as

$$nEF = \frac{EFi}{\sum_{i=1}^{n} EFi} \quad (6)$$

(2) Calculate the average value and percentage as

$$\text{ESA Index} = \frac{nCS+nSNIP+nSJR+nHI+nIF+nEF}{\#features} \times 100 \quad (7)$$

The prefix n represents the normalized value of the indicator, the subscript i represents the ith value of the journal, and the features represent the bibliometric indicators. Figure 5 shows the mean values of the various bibliometric indicators used in this study.




| | CiteScore | SNIP | SJR | H index | Impact Factor | Eigenfactor Score | ESA Index |
|---|---|---|---|---|---|---|---|
| Artificial Intelligence | 5.8959 | 1.6452 | 1.1009 | 49.424 | 2.546 | 0.0035 | 0.4377 |
| Computer Vision and Pattern Recognition | 5.3635 | 1.3572 | 1.0415 | 56.0135 | 2.5141 | 0.0038 | 1.2589 |

Figure 5. Mean value of various indicators in computer science disciplines

## 3.4. Dimensionality reduction

Dimensionality reduction addresses the problem of distinguishing valuable low-dimensional data from high-dimensional data. It represents high-dimensional data as the principal components. In this study, PCA and t-SNE were implemented on a bibliometric dataset to determine the most contributing reduced set of features.

### 3.4.1. Principal component analysis (PCA)

A popular multivariable statistical technique called PCA transformation employs PCA for feature extraction and dimensionality reduction in pattern analysis. By retaining significant information, plainly describing the dataset, and analyzing the observations, it aims to extract significant information from the data and reduce the dataset amount. PCA was employed in this study to reduce the dataset to a new feature space.

### 3.4.2. t-Distributed stochastic neighbor embedding (t-SNE)

The t-SNE algorithm is a novel method of multi-dimensional scaling. This technique is popular because it scales high-dimensional data to low-dimensional data. In this study, this technique was applied to data points (journals) that convert high-dimensional Euclidean distances between data points (journals) into conditional probabilities that represent similarities among journals.

## 3.5. Un-supervised evaluation models

### 3.5.1. Cluster analysis

To identify disconnected groupings in the collected dataset, we used an unsupervised machine learning technique known as k-means clustering. Using an unlabeled dataset, k-means clustering was used to group similar journals. The dataset was divided into groups using the k-means method, and these groups were represented by K variables. In this study, k-means clustering was used to identify clusters based on similar features. Various evaluation measures were used to determine the optimum number of clusters. This verified the percentage of variance as a function of the number of clusters. Based on the pre-evaluated cluster number, the journals were grouped into various numbers of clusters using Euclidean distance.

### 3.5.2. Clustering performance evaluation measures

A clustering algorithm helps to categorize the data. The quality of the clustering results can be assessed using various metrics used for the evaluation.

**Internal evaluation measures**
It is feasible to determine the clustering structure quality without access to external data owing to the internal validation methods. The internal measures are based on information from the input data during clustering. Here, rather than using a ground truth label from the external world, we employed the silhouette coefficient (SC) score, Calinski-Harabasz index (CHI), and Davies-Bouldin index (BDI) for internal cluster validation to assess the cluster quality.

**Determining the optimal value of K**
The elbow method was applied to determine the optimal number of clusters. It examines how the number of groups affects the proportion of the explained variation. The proportion of variation explained by clusters is plotted against the number of clusters. The first clusters would contribute a substantial amount of information. However, eventually the marginal gain would reduce and the graph would adopt an angle. The cluster nodes begin the calculations based on predetermined cluster numbers and are split into clusters based on the predetermined value. The Euclidean distance is used to group the cluster elements into a predetermined number of clusters.

**Silhoutte coefficient (SC) score**
The SC assessment metric was used to assess clustering outcomes. The dissimilarity of a data point or node from other cluster members as well as its similarity to all other points or nodes within its cluster were verified using this clustering validation measure. The SC value lies within [-1, 1]. A higher SC value denotes effective clustering, whereas values near 0 or -1 denote ineffective clustering.

**Calinski-Harabasz index (CHI)**
The CHI is a measure of cluster validity. It is used to evaluate clustering quality. The index is based on the technique used to determine the ratio of between and within-cluster variances. It measures the separation between clusters and their compactness. A higher index value indicates better clustering results.

**Davies-Bouldin index (DBI)**
Davies-Bouldin index (DBI) is used to evaluate the clustering performance. It verifies the inter and intra-cluster similarities of the nodes in clusters based on sample-specific dimensions. The DBI value lies within [0, + ∞]. A value closer to zero indicates a better clustering.

**External evaluation measures**
In the cluster validation process, the external ground-truth label is an additional piece of information incorporated via the external validation approach. When external data are available and there are few true labels in the dataset, an external technique can be used. The effectiveness of the clustering observations was assessed using externally provided data through external validation metrics. In this study, several external validation metrics were used to evaluate the clustering results using available external ground truth data.

**Adjusted Rand score (ARI)**
The adjusted Rand index (ARI) is an external clustering performance evaluation measure. It was used to validate the clustering results with external ground truth labels. In this section, the Scopus and SJR Best Quartiles are used as external class labels for comparison with the clustering labels. The lowest and highest possible values of ARI are -1 and 1, respectively.

**Adjusted mutual information (AMI) score**
The AMI score is a measure of the similarity between two clusters in a dataset. It considers the fact that the mutual information score which measures the amount of information shared by two clusterings, can be biased toward clustering with many small clusters. The AMI score is used here for clustering comparison. The value of Adjusted mutual information ranges from 0 to 1. the value 0 implies dissimilarity and 1 implies most similar clusters.

**Homogeneity, completeness, and V-measure (HCV)**
The homogeneity measures the purity of each cluster with respect to a single class. A clustering result satisfies homogeneity if all its clusters contain only data points that are members of a single class. The homogeneity score ranges from zero to one, with one indicating perfect homogeneity. The completeness measures the extent to which a class is represented by a single cluster. A clustering result satisfies completeness if all the data points that are members of a given class are assigned to the same cluster. The completeness score ranges from zero to one, with one indicating perfect completeness. The V-measure is the harmonic mean of homogeneity and completeness. It provides a single score that balances both the measures. The V-measure score ranges from zero to one, with one indicating perfect agreement between the clustering and true labels. The V-measure is a commonly used metric for clustering evaluations because it considers both homogeneity and completeness.

**Fowlkes-Mallows (FM) score**
The FM score is a measure of the similarity between two clusters in a dataset. This approach is based on the concepts of precision and recall. The score ranges from zero to one, with one indicating perfect agreement between the two clusters, and zero indicating no agreement beyond chance.

**Cross tabulation**

Cross-tabulation places categorical data in a table and then summarizes it by aligning the labels of two classes/categories with each other. Each column of the table contains the number of data members of a class belonging to the data members of another class. It can determine the frequency (either in a raw number or in proportional form) of the values that fall into the groups that the cell is planned to illustrate. Many statistical tests (the majority of which adhere to the chi-squared distribution) can then be performed using the summary data displayed in a cross-tabulated form. In this study, cross-tabulation was used to compare the *Scopus* Quartiles, *SJR Best* Quartiles, and *HJRS* categories with various journal categories observed in the proposed framework.

## 4. Results and discussions

Various impact indicators such as the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor were combined to develop a multi-metric indicator called ESA index for ranking journals. We developed and utilized the index to identify the effects of multiple features using various machine learning techniques. The *Python* libraries *Scikit-learn*, *Matplotlib*, and *Seaborn* were used for these experiments. The experiments were performed using an Intel® Core TM i5 Intel(R) Core (TM) i5-5200U CPU @ 2.20 GHz 2.20 GHz.

### 4.1. Correlation of ESA index with other bibliometric indices

To analyze the ESA index, Spearman's correlation between various bibliometric indicators (i.e., the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal IF) was calculated. Table 3 presents the correlation of artificial intelligence journals. It shows that the ESA index has the highest correlation with the SJR, and a higher correlation with the CiteScore than with the other bibliometric indicators. Table 4 presents the correlation of computer vision and pattern recognition journals. It shows that the ESA index has the highest correlation with the CiteScore, and a higher correlation with the SJR than the other bibliometric indicators. A strong correlation is observed between the ESA index and various bibliometric indicators.

Table 3. Spearman rank correlation between various bibliometric indicators (Artificial Intelligence)

| | CS | SNIP | SJR | H-index | IF | EF | ESA Index |
|---|---|---|---|---|---|---|---|
| **CS** | 1 | | | | | | |
| **SNIP** | 0.86 | 1 | | | | | |
| **SJR** | 0.93 | 0.92 | 1 | | | | |
| **H-index** | 0.66 | 0.56 | 0.63 | 1 | | | |
| **IF** | 0.71 | 0.62 | 0.68 | 0.75 | 1 | | |
| **EF** | 0.66 | 0.59 | 0.66 | 0.80 | 0.96 | 1 | |
| **ESA Index** | 0.92 | 0.88 | 0.93 | 0.81 | 0.84 | 0.84 | 1 |

Table 4. Spearman rank correlation between various bibliometric indicators (Computer Vision and Pattern Recognition)

| | CS | SNIP | SJR | H-index | IF | EF | ESA Index |
|---|---|---|---|---|---|---|---|
| **CS** | 1 | | | | | | |
| **SNIP** | 0.92 | 1 | | | | | |
| **SJR** | 0.97 | 0.96 | 1 | | | | |
| **H-index** | 0.68 | 0.58 | 0.64 | 1 | | | |
| **IF** | 0.74 | 0.69 | 0.73 | 0.65 | 1 | | |
| **EF** | 0.68 | 0.62 | 0.68 | 0.71 | 0.96 | 1 | |
| **ESA Index** | 0.96 | 0.91 | 0.95 | 0.78 | 0.84 | 0.82 | 1 |

*CS, CiteScore; SNIP, Source Normalized Impact per Publication; SJR, SCImago Journal Rank; H-index, Hirsh index; IF, Impact Factor; EF, Eigenfactor; ESA Index, Extended Standardized Average Index

### 4.2. Data analysis of ESA index with benchmark journal quartiles/categories

Table 5 presents a comparison of SA quartiles with *Scopus* quartiles, *SJR Best* quartiles and *HJRS* categories for artificial intelligence journals, and Table 6 shows that of ESA quartiles for artificial intelligence journals using the ARI, AMI score, homogeneity, completeness, V-measure (HCV), and FM score. It shows that various evaluation metrics (while comparing different quartiles) show better results for the ESA index than for the SA index. It can be observed that the *HJRS* has a higher evaluation measure value than the *SJR* and *Scopus* Quartiles. The comparison results of SA quartiles and the ESA quartiles for computer vision and pattern recognition subject area are presented in Tables 7 and 8, respectively. Tables 9 and 10 present the cross-tabulation results for the journals of artificial intelligence and those of computer vision and pattern recognition, respectively.

Table 5. Comparison of SA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Artificial Intelligence)

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus* Quartiles | 0.1755 | 0.1781 | 0.1897 | 0.4394 |
| *SJR* Best Quartiles | 0.1925 | 0.2312 | 0.2282 | 0.4199 |
| *HJRS* Category | 0.3951 | 0.3992 | 0.4261 | 0.5915 |

Table 6. Comparison of ESA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Artificial Intelligence)

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus* Quartiles | 0.3059 | 0.3298 | 0.3446 | 0.5225 |
| *SJR* Best Quartiles | 0.3674 | 0.4630 | 0.4513 | 0.5412 |
| *HJRS* Category | 0.6081 | 0.6164 | 0.6689 | 0.7329 |

Table 7. Comparison of SA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Computer Vision and Pattern Recognition)

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus* Quartiles | 0.1949 | 0.2125 | 0.2521 | 0.4340 |
| *SJR* Best Quartiles | 0.2488 | 0.2985 | 0.3190 | 0.4501 |
| *HJRS* Category | 0.4202 | 0.4338 | 0.4879 | 0.6004 |

Table 8. Comparison of ESA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Computer Vision and Pattern Recognition)

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus* Quartiles | 0.2712 | 0.3265 | 0.3587 | 0.4876 |
| *SJR* Best Quartiles | 0.3599 | 0.4881 | 0.4902 | 0.5320 |
| *HJRS* Category | 0.6138 | 0.6226 | 0.6822 | 0.7343 |

Table 9. Cross tabulations of original journal categories and ESA Index Quartiles in Artificial Intelligence Journals

| | | ESA-Q1 | ESA-Q2 | ESA-Q3 | ESA-Q4 | Total |
|---|---|---|---|---|---|---|
| ***Scopus*-Q** | **Q1** | 71 | 36 | 7 | 0 | 114 |
| | **Q2** | 7 | 25 | 26 | 1 | 59 |
| | **Q3** | 1 | 12 | 27 | 3 | 43 |
| | **Q4** | 0 | 1 | 16 | 18 | 35 |
| ***SJR*-Q** | **Q1** | 55 | 5 | 0 | 0 | 60 |
| | **Q2** | 23 | 32 | 4 | 0 | 59 |
| | **Q3** | 0 | 33 | 27 | 0 | 60 |
| | **Q4** | 0 | 1 | 34 | 21 | 56 |
| **HJRS** | **-** | 0 | 1 | 0 | 3 | 4 |
| | **W** | 70 | 8 | 0 | 0 | 78 |
| | **X** | 8 | 63 | 11 | 0 | 82 |
| | **Y** | 0 | 5 | 58 | 37 | 100 |

Table 10. Cross tabulations of original journal categories and ESA Index Quartiles in Computer Vision and Pattern Recognition Journals

| | | ESA-Q1 | ESA-Q2 | ESA-Q3 | ESA-Q4 | Total |
|---|---|---|---|---|---|---|
| ***Scopus*-Q** | **Q1** | 22 | 12 | 2 | 0 | 36 |
| | **Q2** | 3 | 9 | 6 | 0 | 18 |
| | **Q3** | 0 | 4 | 11 | 1 | 16 |
| | **Q4** | 0 | 1 | 7 | 8 | 16 |
| ***SJR*-Q** | **Q1** | 18 | 3 | 0 | 0 | 21 |
| | **Q2** | 8 | 12 | 0 | 0 | 20 |
| | **Q3** | 0 | 9 | 12 | 0 | 21 |
| | **Q4** | 0 | 0 | 12 | 7 | 19 |
| **HJRS** | **-** | 0 | 1 | 0 | 3 | 4 |
| | **W** | 23 | 5 | 0 | 0 | 28 |
| | **X** | 3 | 18 | 2 | 0 | 23 |
| | **Y** | 0 | 1 | 22 | 14 | 37 |

**Journal categorization using ESA index**

As a comprehensive structure for journal categorization using the ESA index, after calculating this index from six bibliometric indicators, we applied k-means clustering to the dataset with seven features: CiteScore, SNIP, SJR, H-index, Eigenfactor Score, Journal IF, and ESA index. K-means clustering was applied to the full dataset with the seven features. Furthermore, a reduced set of features was obtained through PCA and t-SNE dimensionality reduction techniques. In this section, we demonstrate the experimental observations obtained using the datasets from two aspects. First, we analyze the effect of multiple bibliometric features (the seven features) and with the reduced set of features.

**Clustering results on dataset with seven bibliometric features**

K-means clustering was performed on the dataset for *k* ranging from 2 to 15. Different k values were evaluated because the number of clusters were unknown. For each cluster, various cluster evaluation metrics including the Silhouette Coefficient Score, Calinski-Harabasz score, and Davies-Bouldin Index was computed. This enabled the determination of the value of k at which most cluster validity indices provide the best results. Figure 6 shows the elbow method, Silhouette Coefficient Score is represented in Figure 7, the Calinski-Harabasz score is shown in Figure 8, and the Davies-Bouldin Index is represented in Figure 9. This helps us determine the optimal number of clusters and internal clustering validation results. We selected four clusters for comparison with the *Scopus* Quartiles and *SCImago* Best Quartiles (Q1-Q4). Various experiments were conducted using different k values. Figure 10 presents the k-means clustering results for (a) artificial intelligence and (b) computer vision and pattern recognition.

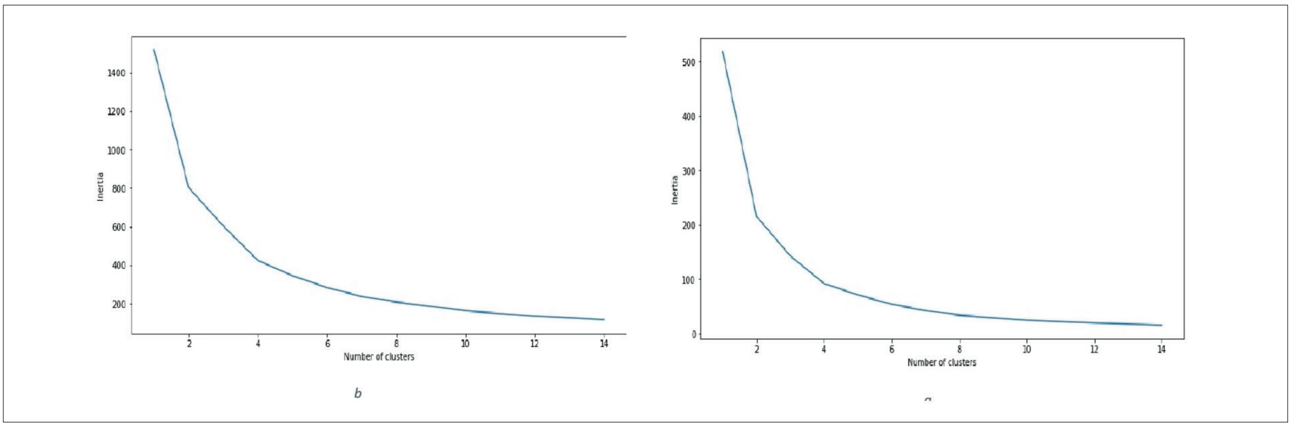



Figure 6. Elbow method (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

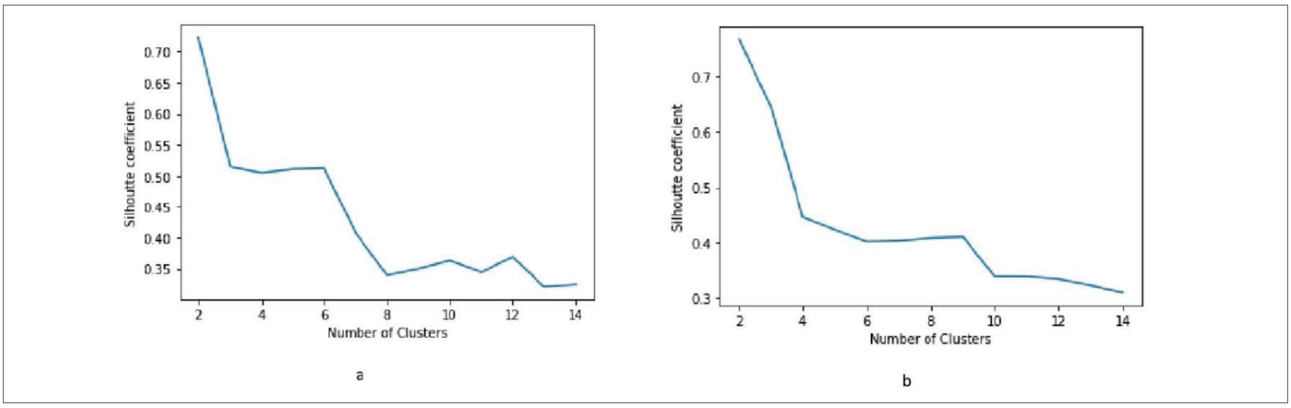



Figure 7 Silhouette Coefficient Score (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

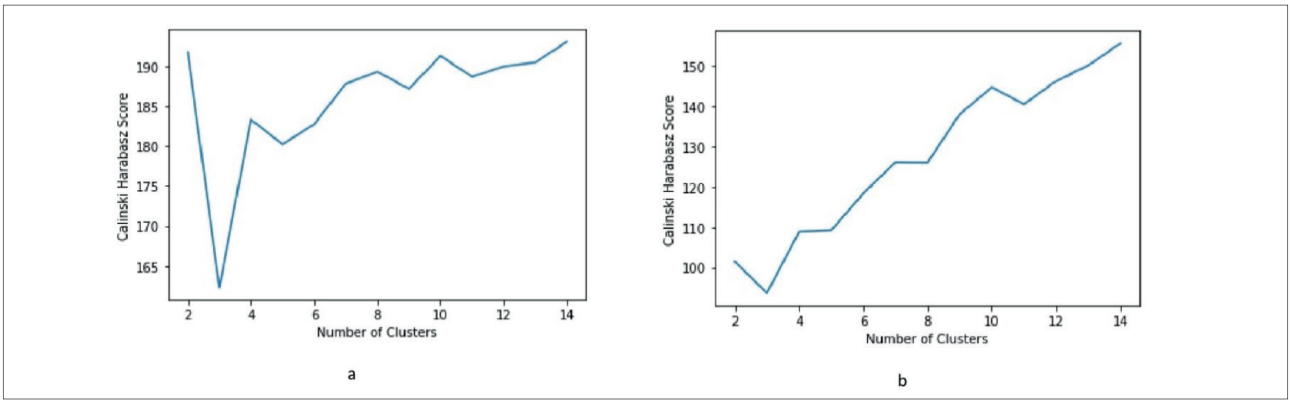



Figure 8. Calinski-Harabasz Score (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

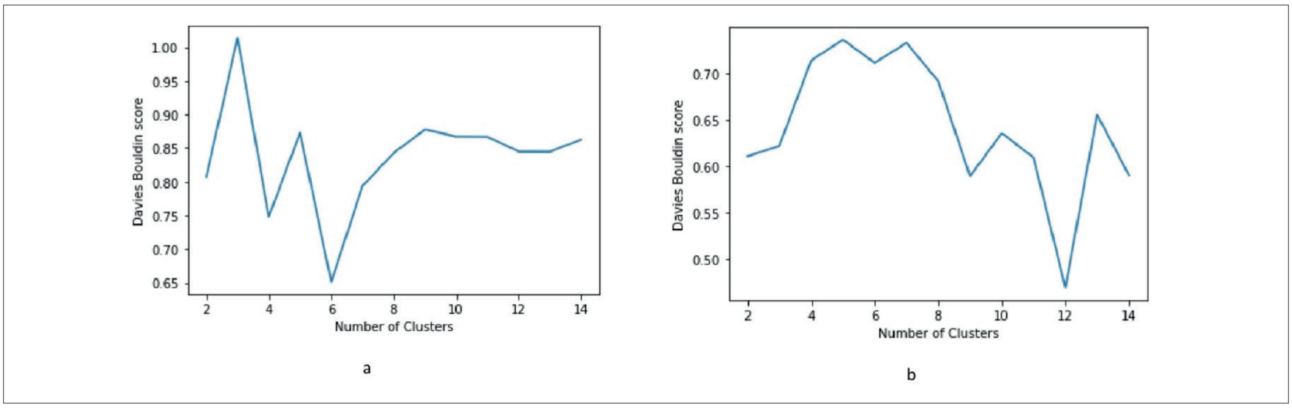


Figure 9. Davies Bouldin method (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

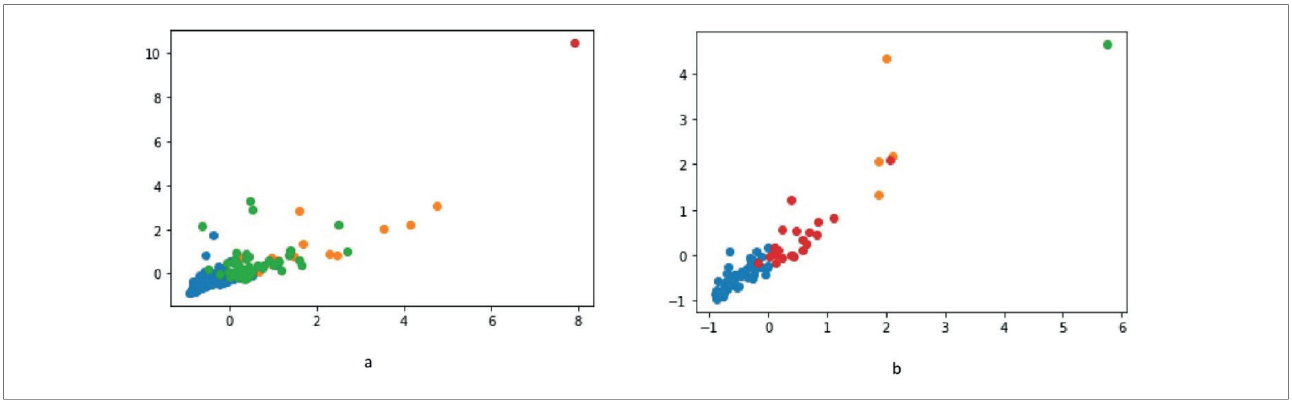


Figure 10. k-means clustering when k = 4 (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

Various clustering validity indices were calculated by applying the seven features as input variables. The *Scopus* Quartiles and *SJR* Best Quartiles were used as target labels. Therefore, in this case, four clusters were selected because the proposed experimental results could be compared with *Scopus* Quartiles and *SJR* Best Quartiles.

### 4.2.1. Comparison of ESA index with the existing benchmarks with all the features

The results of the proposed model were compared with the quartiles of journals from *Scopus* and *SJR*, which are the available standards worldwide. Different evaluation measures were used to measure journal performance.

**Internal evaluation**

The *Scopus* Quartiles and *SJR* Best Quartiles (*SCImago Journal & Country Rank*) provided in this study were used as benchmarks. *Scopus* categorize journals into four quartiles Q1-Q4. Here, Q1 is the top-ranking group. It is followed by Q2, Q3 and Q4 which is the lowest category. The *SCImago Journal & Country Rank* also classifies journals into four categories: Q1-Q4. The categories of each journal used as the input dataset were obtained from *Scopus* and *SJR*.

*Scopus* and *SJR* are two existing systems for journal categorization and rating. The current strategy differs primarily in that *Scopus* is based on set standards developed by certain statistical measures based on a single metric (i.e., it is calculated on the basis of the CiteScore, and *SCImago Journal & Country Ranks* use SJR for journal categorization). SJR is based on generic frameworks that learn automatically from data. In this experiment, six baseline features (i.e., the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor) were used as input features. In addition, a new feature known as the ESA index was calculated and used in this analysis. The journal categories obtained in the proposed model were validated using different internal evaluation metrics. The results of this experiment are shown in Table 11 for artificial intelligence journals and in Table 12 for computer vision and pattern recognition journals.

Table 11. Internal clustering validity results for all features n = 7 (Artificial Intelligence)

| No. of clusters | Silhouette score | Calinski-Harabasz score | Davies-Bouldin Index |
|---|---|---|---|
| K = 2 | 0.7103 | 191.7727 | 0.8072 |
| K = 3 | 0.5154 | 162.3236 | 1.0136 |
| K = 4 | 0.5052 | 183.3192 | 0.7478 |
| K = 5 | 0.5112 | 180.2280 | 0.8713 |
| K = 6 | 0.5127 | 184.2999 | 0.6512 |
| K = 7 | 0.3831 | 187.6632 | 0.7859 |

Table 12. Internal clustering validity results for all features n = 7 (Computer Vision and Pattern Recognition)

| No. of clusters | Silhouette score | Calinski-Harabasz score | Davies-Bouldin Index |
|---|---|---|---|
| K = 2 | 0.7658 | 101.5086 | 0.6111 |
| K = 3 | 0.6852 | 87.9538 | 0.5589 |
| K = 4 | 0.4457 | 108.7583 | 0.7142 |
| K = 5 | 0.4226 | 109.2883 | 0.7366 |
| K = 6 | 0.4015 | 118.3215 | 0.7175 |
| K = 7 | 0.4071 | 126.0863 | 0.7317 |

**External evaluation**

Tables 13 and 14 present the k-means clustering validation results compared with the *Scopus* Quartiles and *SJR* Best Quartiles as ground-truth labels using different evaluation metrics for (1) artificial intelligence and (2) computer vision and pattern recognition, respectively. The *SJR* Quartiles showed relatively better results than the *Scopus* Quartiles.

Table 13. External validation of clustering labels with *Scopus* and *SCImago Journal Rank* (Artificial Intelligence) journals

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus*-Q | 0.0329 | 0.2450 | 0.2147 | 0.4373 |
| *SJR*-Q | 0.1754 | 0.3313 | 0.2724 | 0.4892 |

Table 14. External validation of clustering labels with *Scopus* and *SCImago Journal Rank* (Computer Vision and Computer Vision) journals

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus*-Q | 0.1132 | 0.2549 | 0.2526 | 0.4501 |
| *SJR*-Q | 0.2642 | 0.4210 | 0.3728 | 0.5277 |

## 4.3. Clustering results on the reduced dataset using PCA

Seven bibliometric indicators were selected as inputs for categorizing journals through k-means clustering. Then, PCA was applied. It transformed the seven-dimensional dataset into seven PCs. The variance explained by each PC based on the input dataset demonstrates that PCA can be used successfully in the categorization of journal datasets for dimensionality reduction because the first two PCs maintained approximately 89% of the variation for artificial intelligence journals and 94% for computer vision and pattern recognition journals.

Table 15. Internal clustering validity results for the reduced set of features PC1 and PC2 (Artificial Intelligence)

| No. of clusters | Silhouette score | Calinski-Harabasz score | Davies Bouldin Index |
|---|---|---|---|
| K = 2 | 0.7366 | 241.5159 | 0.6785 |
| K = 3 | 0.5679 | 217.2935 | 0.8123 |
| K = 4 | 0.5641 | 271.5580 | 0.5888 |
| K = 5 | 0.5592 | 293.3024 | 0.5085 |
| K = 6 | 0.5601 | 305.7263 | 0.5595 |
| K = 7 | 0.4639 | 338.2717 | 0.6228 |

Table 16. Internal clustering validity results for the reduced set of features PC1 and PC2 (Computer Vision and Pattern Recognition)

| No. of clusters | Silhouette score | Calinski-Harabasz score | Davies Bouldin Index |
|---|---|---|---|
| K = 2 | 0.7874 | 117.7504 | 0.5591 |
| K = 3 | 0.6772 | 118.1159 | 0.5432 |
| K = 4 | 0.5362 | 154.2243 | 0.6043 |
| K = 5 | 0.4805 | 167.1288 | 0.6620 |
| K = 6 | 0.4836 | 202.6179 | 0.5595 |
| K = 7 | 0.4950 | 250.6245 | 0.5217 |

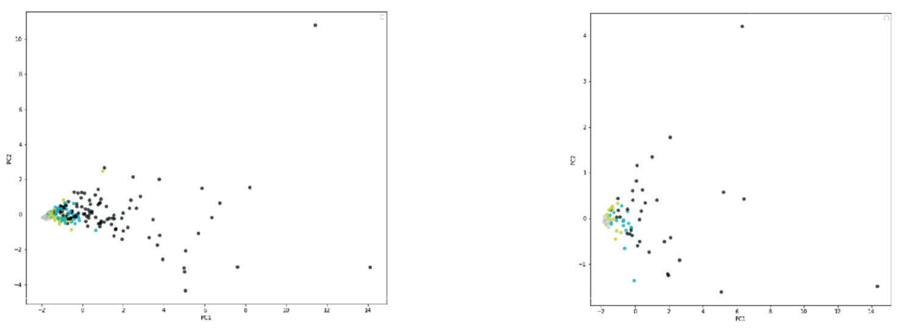
We projected the k-means derived clusters onto 2D visuals after applying PCA to divide the dataset into two principal components. For the PCA, k-means was applied. By processing a lower-dimensional dataset via k-means, the score value increased from 0.50 to 0.56 (see Tables 15 and 16). A significant improvement in the capability to distinguish between clusters is observed in the 2D scatter plots. Table 17 and 18 display the external validity scores for artificial intelligence journals and computer vision and pattern recognition journals, respectively, when the *Scopus* and *SJR* Quartiles are employed as ground truth labels.

Table 17. External validation of PCA clustering labels with *Scopus* and *SCImago Journal Rank* (Artificial Intelligence) journals

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus*-Q | 0.0329 | 0.2450 | 0.2147 | 0.4373 |
| *SJR*-Q | 0.1850 | 0.3427 | 0.2812 | 0.4959 |

Table 18 External validation of PCA clustering labels with *Scopus* and *SCImago Journal Rank* (Computer Vision and Pattern Recognition) journals

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus*-Q | 0.0266 | 0.2523 | 0.2428 | 0.4217 |
| *SJR*-Q | 0.2104 | 0.3610 | 0.3160 | 0.5146 |



a b

Figure 11. K-means clustering with PCA for k= 4 (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

### 4.4. Clustering results on the reduced dataset using t-SNE

In this section, we reduced our dataset using t-SNE and compared the k-means results with those of the PCA k-means. The dataset was reduced to two t-SNE components. The data tended to cluster into a large diffused cluster with a perplexity of 80 for artificial intelligence journals and 70 for computer vision and pattern recognition journals.

The Silhouette Coefficient Score, that we achieved by applying k-means to our two t-SNE-derived components was 0.42, whereas that we acquired by applying k-means to the two principal components of PCA was 0.56. The interpretation of t-SNE appears counterintuitive because the density of t-SNE clusters (i.e. low-dimensional space) is not proportionally related to data associations in the original (high-dimensional space) dataset. That is, although we can have good dense clusters generated by k-means, t-SNE may reveal these as broad or even numerous clusters. This is particularly so when the perplexity is excessively low. When interpreting the t-SNE plots, it is difficult to interpret the density, cluster size, number of clusters (under the same k-means cluster), and form. Although we can have numerous clusters for the same k-means cluster (particularly when the perplexity is significantly low), this has no bearing on the cluster quality. The distance and location of each k-means cluster are the key advantages of t-SNE. Although clusters that are closer together are more closely related to each other, this does not necessarily imply that clusters that are farther apart are proportionally dissimilar. Finally, we need to observe a particular level of separation between the k-means clusters, as shown by t-SNE.

Table 19. Internal clustering validity results for the reduced set of features t-SNE1 and t-SNE2 (Artificial Intelligence)

| No. of clusters | Silhouette score | Calinski-Harabasz score | Davies-Bouldin Index |
|---|---|---|---|
| K = 2 | 0.5544 | 477.6020 | 0.6155 |
| K = 3 | 0.4946 | 510.4948 | 0.6964 |
| K = 4 | 0.4201 | 470.9313 | 0.8121 |
| K = 5 | 0.4267 | 481.5732 | 0.8063 |
| K = 6 | 0.4359 | 480.2519 | 0.7802 |
| K = 7 | 0.4433 | 491.9881 | 0.7691 |

Table 20. Internal clustering validity results for the reduced set of features t-SNE1 and t-SNE2 (Computer Vision and Pattern Recognition)

| No. of clusters | Silhouette score | Calinski-Harabasz score | Davies-Bouldin Index |
|---|---|---|---|
| K = 2 | 0.6244 | 118.4537 | 0.6302 |
| K = 3 | 0.4878 | 135.8395 | 0.6597 |
| K = 4 | 0.4794 | 151.3629 | 0.7169 |
| K = 5 | 0.4968 | 162.5077 | 0.6577 |
| K = 6 | 0.4944 | 173.9182 | 0.6564 |
| K = 7 | 0.4984 | 172.9062 | 0.5650 |

We projected the k-means-derived clusters onto 2D visuals after using t-SNE to divide the dataset into two components. k-means was applied to t-SNE. By processing a lower-dimensional dataset via k-means, the score value increased from 0.50 to 0.48. This is shown in Table 19 for artificial intelligence and Table 20 shows results for computer vision and pattern recognition. A significant improvement in the capability to distinguish between clusters is observed in the 2D scatter plots. Tables 21 and 22 display the external validity scores when the *Scopus* and *SJR* Quartiles are employed as ground truth labels. Tables 21 and 22 display the external validity scores for artificial intelligence, and computer vision and pattern recognition respectively.

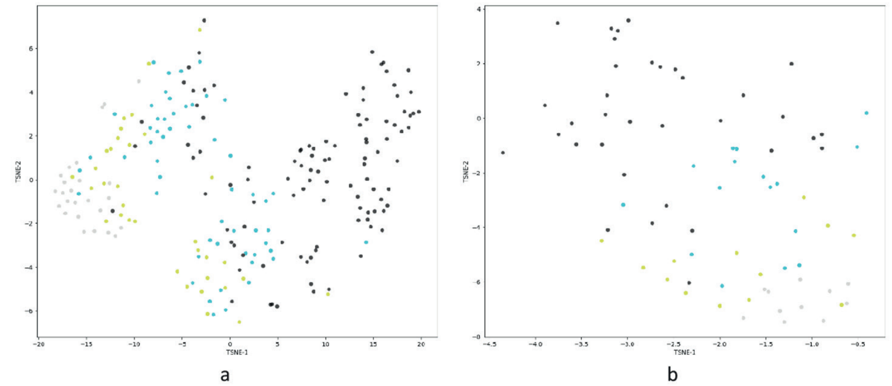



Figure 12. K-means clustering visualization with t-SNE (a) Artificial Intelligence (b) Computer vision and pattern recognition

Table 21. External validation of t-SNE clustering labels with *Scopus* and *SCImago Journal Rank* (Artificial Intelligence) journals

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus*-Q | 0.2391 | 0.3342 | 0.3655 | 0.4574 |
| *SJR*-Q | 0.3699 | 0.4331 | 0.4427 | 0.5265 |

Table 22. External validation of t-SNE clustering labels with *Scopus* and *SCImago Journal Rank* (Computer Vision and Computer Vision) journals

| Quartiles | ARI | MI | HCV | FM |
|---|---|---|---|---|
| *Scopus*-Q | 0.1032 | 0.2746 | 0.3021 | 0.3927 |
| *SJR*-Q | 0.3517 | 0.4764 | 0.4655 | 0.5448 |

## 5. Conclusions and future work

Various researchers have disapproved the evaluation of the scientific impact of a journal using an individual indicator such as the Journal Impact Factor (IF). Furthermore, the Journal IF is not only widely applied but also often misapplied. This has yielded biased and misleading results. Earlier, the HEC used the IF for research evaluations. Owing to the bias induced by individual indicators, a multi-metric journal prestige measurement system is necessary for journal quality estimation. In Pakistan, the *HEC Journal Recognition System* (*HJRS*) was launched in July 2020 to evaluate journals using proprietary JPI measures that divide journals into the W, X, and Y categories.
*https://HJRS.hec.gov.pk*

The *HJRS* is a multi-metric tool used to categorize journals based on the Eigenfactor Score, Article Influence (AI) Score, SCImago Journal Rank (SJR), SNIP, CD2, and H-index. However, it has few limitations: (1) CiteScore is a well-known journal-based metric launched by *Elsevier* (*Scopus*). It directly competes with the Journal Impact Factor (IF). It has not been used in the *HJRS* for journal categorization. (2) According to a few researchers, the decision-making mechanism of the *HJRS* is not satisfactory. This is because few journals that have been reported earlier in the W category have now been shifted to lower categories in the *HJRS*. This increases conflict rather than facilitating research in Pakistan. (3) Many researchers consider that HEC should reduce the threshold levels for certain categories. Therefore, the proposed study attempts to address these issues of *HJRS*. In this regard, a multi-metric-based approach was adopted from the SA index, which used two bibliometric measures: the IF and H-index. In this study, a multi-metric-based extended standardized average (ESA) index was developed using six metrics: CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor from three databases (*Scopus*, *SCImago Journal & Country Rank*, and *Web of Science*). The CiteScore was included to overcome the first issue of the *HJRS*. Second, the proposed model is not based on proprietary measures that makes the system transparent. The ESA index is strongly correlated with other well-known bibliometric indicators. Thus, this framework enhances the overall efficiency of journal ranking systems by aggregating multiple bibliometric indicators. The ESA index performed better than the SA Index and was highly correlated with all the other bibliometric indicators. Furthermore, a machine-learning based evaluation was performed on the proposed study to determine the combined impact of the ESA index with other metrics. In addition, k-means clustering coupled with dimensionality reduction techniques such as PCA and t-SNE was applied to identify hidden patterns in journal categorization. The proposed model examined the effectiveness of the journal prestige measurement system for all seven features and a reduced set of features. Based on the clustering evaluation measure and world benchmark bibliometric indices, we selected the optimum number of clusters as $k = 4$ (which indicated four clusters). The proposed model results were compared with the *Scopus* and *SCImago* Best Quartiles (Q1-Q4) and the *HJRS* Categories (W, X, and Y) using cross-tabulation. The results showed that compared with the use of the seven features for journal categorization, reduced/transformed features provided superior results with dimensionality reduction techniques such as PCA and t-SNE. It is concluded that the multi-metric ESA index can be used to facilitate the decision-making process with regard to the selection of venues for publishing research articles. Furthermore, the use of this index can also assist in predicting future performance of the selected journals.

There are several approaches to expand the scope of this study. It concentrated on computer science journals to construct the dataset. A convenient expansion would be to develop a dataset of other areas and subfields in the computer science domain (such as software, data communication and networks) and then, apply the proposed system to a new dataset to identify the patterns in other subjects. Furthermore, clustering, feature selection, and classification techniques can be used to further evaluate the framework. Other prestigious journal rankings can be used to compare the results. This would facilitate the examination of the patterns of journal popularity or decline over time.

## 6. Statements and declarations

The data collection, experimentation, and initial draft writing was carried out by the first author. The second author suggested the concept for the article, aided in analyzing the results, and revised the initial draft. The third author enhanced the experimental design and validated the authenticity of the experiments. The fourth author contributed to the refinement of the writing, organization of the concepts, research coordination, and professional editing.

## 7. References

**Abbas, Naveed-Neem**; **Ahmed, Tanveer**; **Shah, Syed-Habib-Ullah**; **Omar, Muhammad**; **Park, Han Woo** (2019). "Investigating the applications of artificial intelligence in cyber security". *Scientometrics*, v. 121, n. 2, pp. 1189-1211.
*https://doi.org/10.1007/s11192-019-03222-9*

**Allen, Liz**; **Jones, Ceri**; **Dolby, Kevin**; **Lynn, David**; **Walport, Mark** (2009). "Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs". *PLoS one*, v. 4, n. 6.
*https://doi.org/10.1371/journal.pone.0005910*

**Beliakov, Gleb**; **James, Simon** (2011). "Citation-based journal ranks: The use of fuzzy measures". *Fuzzy sets and systems*, v. 167, n. 1, pp. 101-119.
*https://doi.org/10.1016/j.fss.2010.08.011*

**Bergstrom, Carl** (2007). "Eigenfactor: Measuring the value and prestige of scholarly journals". *College & research libraries news*, v. 68, n. 5, pp. 314-316.
*https://doi.org/10.5860/crln.68.5.7804*

**Bollen, Johan**; **Van-de-Sompel, Herbert**; **Hagberg, Aric**; **Chute, Ryan** (2009). "A principal component analysis of 39 scientific impact measures". *PLoS one*, v. 4, n. 6.
*https://doi.org/10.1371/journal.pone.0006022*

**Dellavalle, Robert P.**; **Schilling, Lisa M.**; **Rodriguez, Marko A.**; **Van-de-Sompel, Herbert**; **Bollen, Johan** (2007). "Refining dermatology journal impact factors using PageRank". *Journal of the American Academy of Dermatology*, v. 57, n. 1, pp. 116-119.
*https://doi.org/10.1016/j.jaad.2007.03.005*

**Duan, Lian**; **Nasri, Farrokh**; **Paknejad, Javad** (2018). "A data-driven method for rating management information systems journals in the same scale of the Association of Business Schools Journal Guide". *Expert systems*, v. 35, n. 6.
*https://doi.org/10.1111/exsy.12309*

**Ennas, Gianfranco**; **Biggio, Battista**; **Di-Guardo, Maria-Chiara** (2015). "Data-driven journal meta-ranking in business and management". *Scientometrics*, v. 105, n. 3, pp. 1911-1929.
*https://doi.org/10.1007/s11192-015-1751-y*

**Feng, Lin**; **Zhou, Jian**; **Liu, Sheng-Lan**; **Cai, Ning**; **Yang, Jie** (2020). "Analysis of journal evaluation indicators : an experimental study based on unsupervised Laplacian score". *Scientometrics*, v. 124, pp. 233-254.
*https://doi.org/10.1007/s11192-020-03422-8*

**Fernández-Cano, Antonio**; **Fernández-Guerrero, Inés M.** (2017). "A multivariate model for evaluating emergency medicine journals". *Scientometrics*, v. 110, n. 2, pp. 991-1003.
*https://doi.org/10.1007/s11192-016-2197-6*

**Ferrer-Sapena, Antonia**; **Sánchez-Pérez, Enrique A.**; **Peset, Fernanda**; **González, Luis-Millán**; **Aleixandre-Benavent, Rafael** (2016). "The impact factor as a measuring tool of the prestige of the journals in research assessment in mathematics". *Research evaluation*, v. 25, n. 3, pp. 306-314.
*https://doi.org/10.1093/reseval/rvv041*
*https://academic.oup.com/rev/article-abstract/25/3/306/2364555*

**Fersht, Alan** (2009). "The most influential journals: Impact factor and eigenfactor". *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, n. 17, pp. 6883-6884.
*https://doi.org/10.1073/pnas.0903307106*

**González-Pereira, Borja**; **Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix** (2010). "A new approach to the metric of journals scientific prestige: The SJR indicator". *Journal of informetrics*, v. 4, n. 3, pp. 379-391.
*https://doi.org/10.1016/j.joi.2010.03.002*

**Halim, Zahid; Khan, Shafaq** (2019). "A data science based framework to categorize academic journals". *Scientometrics*, v. 119, n. 1.
*https://doi.org/10.1007/s11192-019-03035-w*

**Holmberg, Kim**; **Park, Han Woo** (2018). "An altmetric investigation of the online visibility of South Korea-based scientific journals". *Scientometrics*, v. 117, n. 1, pp. 603-613.
*https://doi.org/10.1007/s11192-018-2874-8*

**Hsu, Wen-Chin**; **Tsai, Chih-Fong**; **Li, Jia-Huan** (2015). "A hybrid indicator for journal ranking: An example from the field of Health Care Sciences and Services". *Online information review*, v. 39, n. 7, pp. 858-869.
*https://doi.org/10.1108/OIR-11-2014-0277*

**James, Chris**; **Colledge, Lisa**; **Meester, Wim**; **Azoulay, Norman**; **Plume, Andrew** (2018). "CiteScore metrics: Creating journal metrics from the *Scopus* citation index". *ArXiv*, 16 pp.
*https://onlinelibrary.wiley.com/doi/epdf/10.1002/leap.1246*