Document-type classification errors in bibliometric databases: Insights from the engineering/manufacturing field

Materials Research Proceedings 57 (2025) 80-89 https://doi.org/10.21741/9781644903735-10

Document-type classification errors in bibliometric databases: Insights from the engineering/manufacturing field

Domenico Augusto MAISANO[1,a *], Lucrezia FERRARA[1,b] and Fiorenzo FRANCESCHINI[1,c]

[1] Politecnico di Torino, Department of Management and Production Engineering (DIGEP), Corso Duca degli Abruzzi 24, 10129, Torino (Italy)

[a] domenico.maisano@polito.it, [b] lucrezia.ferrara@polito.it, [c] fiorenzo.franceschini.polito.it

Keywords: Document-Type Classification, Quality, Performance Indicators

Abstract. Document types (DTs) – e.g., *research articles, reviews, conference proceedings, letters*, etc. – are not only used to classify scientific publications, but also to routinely guide inclusion-exclusion decisions in bibliometric assessments, often without adequate consideration of the quality of underlying content. This study examines DT-classification errors in Scopus and Web of Science (WoS), focusing on engineering/manufacturing publications. These errors – which may directly affect publication/citation counts, citation-impact indicators, and consequently academic evaluations and careers – are analyzed in a *corpus* of about 10,000 documents, using a recent semi-automated method. The results indicate that these errors, while occurring in several percentage points, are far from negligible. Furthermore, statistical analyses reveal systematic differences among publishers (e.g., Springer, Elsevier, Taylor & Francis, etc.), with some contributing more to errors, probably due to editorial styles or inconsistent metadata. This study provides insights for researchers, evaluators and database managers, highlighting the need for publisher-specific guidelines to enhance classification accuracy and reduce errors.

1. Introduction

Document types serve as classification labels for scientific documents (e.g., *research articles, reviews, conference proceedings*) and facilitate targeted bibliographic searches. These classifications, assigned by authors, editorial committees, publishers, or bibliometric databases, are not always consistent, leading to discrepancies in document categorization [1,2]. This study examines DT classifications within Scopus and WoS, two leading bibliometric databases widely utilized for research retrieval and evaluation [3]. DT-classification errors in these databases can have significant implications in academic assessments, including research evaluations for individual scholars, research groups, and institutions [4]. Prior studies indicate that such errors, while occurring at relatively low rates, are non-negligible and may lead to document misclassification, affecting research assessments and academic decisions [5,6].

Maisano et al. [7] recently introduced a semi-automated methodology for detecting DT-classification errors by analyzing discrepancies between Scopus and WoS. This approach, which combines automated identification with manual verification, enables large-scale analysis while minimizing the labour-intensive nature of prior studies. Building on this work, the present study focuses on publications from Politecnico di Torino (PoliTO) and addresses three key questions:

***RQ1**: What is the incidence of DT-classification errors in engineering/manufacturing publications?*

***RQ2**: Do publishers systematically influence DT-classification errors?*

***RQ3**: Are error distributions consistent across Scopus and WoS?*

To investigate these questions, a structured methodological framework is adopted, comprising document and publisher identification, data consolidation (i.e., grouping sub-publishers into

macro-publishers or publishing groups to avoid excessive fragmentation in the analysis), the detection of DT-classification errors, statistical analysis, and comparative assessment of DT-classification discrepancies across databases. The remainder of this article is organized as follows: Sect. 2 outlines the methodological framework, Sect. 3 presents the empirical findings, and Sect. 4 the conclusions, practical implications, limitations, and directions for future research.

2. Methodology

This section details the methodological framework adopted in the study, which is structured into two key phases: (a) data collection and identification of DT-classification errors in Scopus and WoS and (b) publisher identification and aggregation into macro-publishers, along with computation of relevant statistics.

2.1 Data collection and analysis of DT-classification errors

Building on previous research [7], this study analyzes a dataset of 10,834 scientific documents authored by PoliTO faculty members from 2019 to 2023. The dataset, indexed in both Scopus and WoS and identified through DOI codes, ensures a consistent and comparable document pool within the field of engineering/manufacturing, a discipline characterized by relatively homogeneous publishing practices [8]. To detect DT-classification errors, a semi-automated methodology [7] was adopted. The initial phase identified 10,471 concordant documents, (i.e., classified identically in both databases), and 363 discordant documents which exhibit classification discrepancies (Table 1). The manual analysis of all 363 discordant documents confirmed a substantial number of classification errors.

Table 1 *– Concordance matrix for the corpus of interest of 10,834 documents. While the elements in the diagonal (in round brackets) denote DT classifications that are concordant between competing databases, those off-diagonal denote potential DT-classification errors. Documents corresponding to the latter elements were manually analysed.*

DT classifications →		by Scopus									
↓		*Article*	*Conf. paper*	*Review*	*Editorial*	*Book chapter*	*Erratum*	*Note*	*Letter*	*Short Survey*	**Row total**
by WoS	*Article*	(8,378)	29	100	3	3	–	10	6	2	**8,531**
	Proceedings Paper	15	(1,463)	–	1	55	–	–	–	–	**1,534**
	Review	70	–	(413)	–	–	–	–	–	3	**486**
	Editorial Material	20	4	17	(162)	–	1	14	–	3	**221**
	Correction	1	–	–	–	–	(48)	–	–	–	**49**
	Letter	2	–	–	–	–	–	–	(7)	–	**9**
	Retraction	–	–	–	–	–	2	–	–	–	**2**
	Meeting Abstract	–	1	–	–	–	–	–	–	–	**1**
	Book Review	1	–	–	–	–	–	–	–	–	**1**
	Column total	**8,487**	**1,497**	**530**	**166**	**58**	**51**	**24**	**13**	**8**	**10,834**

Tables 2 and 3 present error matrices for Scopus and WoS, where rows indicate the "true" DTs (determined through manual validation), and columns reflect the DTs assigned by each database. Diagonal elements denote correct classifications, while off-diagonal entries indicate misclassifications. Common errors include the misclassification of *research articles* as *reviews* or *conference papers* and *vice versa*. The overall error rate (ε), calculated as:

$$\varepsilon = \frac{\sum_{\forall(i,j) \mid i \neq j} x_{(i,j)}}{\sum_{\forall(i,j)} x_{(i,j)}}, \quad (1)$$

where $x_{(i,j)}$ represents the number of documents reported in the i-th row and j-th column of the error table, for the database of interest. The ε values resulting from the analysis are shown in the lower right vertex of the error tables (i.e., 2.1% for Scopus and 1.3% for WoS) in Tables 2 and 3, consistent with prior studies [6,7,9]. These findings confirm that while both databases exhibit classification inconsistencies, the overall error rate remains within a few percentage points.

Table 2 *– Error table for Scopus, with error rate (ε, cf. Eq. 1) highlighted in bold. The diagonal elements indicating correctly classified documents are marked with round brackets.*

		DT classification by Scopus									
		Article	*Conf. Paper*	*Review*	*Editorial*	*Book chapter*	*Erratum*	*Note*	*Letter*	*Short Survey*	Row total
"True" DT classifications	*Article*	(8,400)	23	28	-	3	-	2	-	1	8,457
	Conf. Paper	15	(1,473)	-	1	51	-	-	-	-	1,540
	Review	50	-	(483)	-	-	-	-	-	1	534
	Editorial	14	1	14	(164)	-	-	3	-	3	199
	Erratum	1	-	-	-	-	(49)	-	-	-	50
	Note	5	-	2	1	-	-	(19)	-	1	28
	Letter	1	-	-	-	-	-	-	(13)	-	14
	Book Chapter	-	-	-	-	(4)	-	-	-	-	4
	Other	1	-	3	-	-	-	-	-	-	4
	Short Survey	-	-	-	-	-	-	-	-	(2)	2
	Retracted	-	-	-	-	-	2	-	-	-	2
	Column total	8,487	1,497	530	166	58	51	24	13	8	10,834
											$\varepsilon \cong 2.1\%$

Table 3 *– Error table for WoS, with error rate (ε, cf. Eq. 1) highlighted in bold. The diagonal elements indicating correctly classified documents are marked with round brackets.*

		DT classification WoS									
		Article	*Proc. Paper*	*Review*	*Editorial Material*	*Correction*	*Letter*	*Retraction*	*Book Review*	*Meeting Abstract*	Row total
"True" DT classifications	*Article*	(8,441)	-	17	1	-	1	-	-	-	8,460
	Proc. Paper	6	(1530)	-	3	-	-	-	-	1	1,540
	Review	63	-	(467)	2	-	-	-	-	-	532
	Editorial Material	3	-	-	(198)	-	-	-	-	-	201
	Correction	-	-	-	1	(49)	-	-	-	-	50
	Other	8	-	2	15	-	-	-	-	-	25
	Letter	6	-	-	-	-	(8)	-	-	-	14
	Book Chapter	-	4	-	-	-	-	-	-	-	4
	News Item	3	-	-	-	-	-	-	-	-	3
	Retraction	-	-	-	-	-	-	(2)	-	-	2
	Book Review	1	-	-	-	-	-	-	(1)	-	2
	Biographical-Item	-	-	-	1	-	-	-	-	-	1
	Column total	8,531	1,534	486	221	49	9	2	1	1	10,834
											$\varepsilon \cong 1.3\%$

2.2 Identification and rationalization of publishers

This phase of the analysis examines publishers as the entities responsible for the editorial management of scientific documents, a role distinct from the scientific content itself, which is typically overseen by expert committees and peer reviewers [8]. Publisher information for each document was extracted using the "Publisher" field in Scopus and WoS.

The initial analysis revealed significant fragmentation, with a total of 333 distinct publishers identified across the dataset. This fragmentation arises not only from the existence of genuinely independent publishers but also from the structural complexity of major publishing groups. Large publishers frequently operate under multiple "sub-brands", which bibliometric databases may report inconsistently. For instance, Elsevier is variously listed as "Elsevier GmbH," "Elsevier USA," "Elsevier Masson s.r.l.," and "Elsevier B.V.," among others. Similarly, naming inconsistencies exist within other major publishers, such as variations in the representation of John Wiley & Sons (e.g., "John Wiley and Sons Inc.", "John Wiley and Sons Ltd.").

To address these inconsistencies, a data cleaning and aggregation process was undertaken, consolidating sub-publishers into broader macro-publishers. The rationale for this approach is that macro-publishers impose standardized editorial policies, work practices, and metadata conventions, ensuring a degree of consistency across their sub-brands. Additionally, aggregating sub-publishers reduces dataset fragmentation, increasing the statistical robustness of subsequent analyses [10]. The manual aggregation process relied on online sources detailing publisher affiliations with broader editorial groups, which was one of the most laborious aspects of the entire analysis. Following this rationalization, 218 macro-publishers were identified. To enhance

statistical reliability, only macro-publishers with at least 300 documents were considered in further analyses, resulting in a subset of 7 macro-publishers, with the remaining entities grouped under the category "Other". For simplicity, the term "publishers" will henceforth refer to these 7 macro-publishers. The cumulative chart in Fig. 1 reveals that the top 7 publishers listed account for approximately 80% of the total documents included in the analysis (i.e., $\frac{Documents\ of\ the\ major\ 7\ macro-publishers}{Total\ corpus\ of\ documents} = \frac{8,629}{10,834} \approx 80\%$) and 79% of the identified discordant documents (i.e., $\frac{Discordant\ documents\ of\ the\ major\ 7\ macro-publishers}{Discordant\ documents\ in\ the\ total\ corpus} = \frac{286}{363} \approx 79\%$). This finding reinforces the rationale for excluding macro-publishers with fewer documents from the analysis, grouping them into the "Other" category.

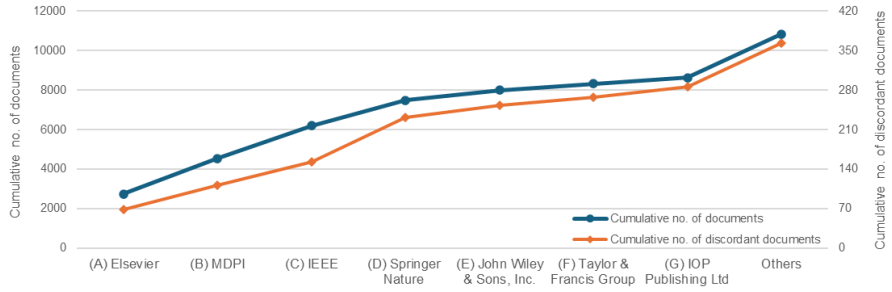


Figure 1 *– Cumulative diagram of the total number of documents considered in the analysis (left-hand scale) and the total number of discordant documents (right-hand scale). The macro-publishers on the x-axis are listed in descending order based on the total number of documents included in the analysis (cf. Table 4).*

3. Results

This section examines the results of the analysis of the decomposition of error tables related to the documents of specific publishers for each database. Analyses are supported by various statistical tests.

The error statistics for documents associated with a given p-th publisher can be modelled as a binomially distributed variable, characterized by the following mean (μ) and variance (σ^2):

$$E_p \sim Bin[\mu = d_p \cdot \varepsilon_p, \sigma^2 = d_p \cdot \varepsilon_p \cdot (1 - \varepsilon_p)], \tag{2}$$

where d_p represents the total number of documents included in the analysis and ε_p is the error rate from the perspective of the p-th publisher.

When the condition $d_p \cdot \varepsilon_p \geq 5$ is satisfied, e_p can be approximated by a normal distribution with the same mean and variance as above[1]:

$$E_p \sim N\left[\mu = d_p \cdot \varepsilon_p, \sigma^2 = d_p \cdot \varepsilon_p \cdot (1 - \varepsilon_p)\right]. \tag{3}$$

Based on this approximation, a symmetric 95% confidence interval (CI) around the estimated error rate $\hat{\varepsilon}_p$ can be defined as follows:

$$\hat{\varepsilon}_p \pm z_{1-\frac{\alpha}{2}} \cdot \sqrt{\frac{\hat{\varepsilon}_p \cdot (1-\hat{\varepsilon}_p)}{d_p}}, \tag{4}$$

where:

α is the type-I error (e.g., α=5% for a 95% CI);

[1] This condition is generally met when considering relatively large sets of misclassified documents (e_p). However, even when it is not strictly met, Eq. 4 still provides a practical and reasonable estimate of the confidence interval in Eq. 4 [10].

$z_{1-\frac{\alpha}{2}}$ is the unit normal deviate corresponding to a cumulative probability $1-\frac{\alpha}{2}$. E.g., for a symmetrical 95% CI, α = 5%, hence $z_{0.975} \approx 1.96$.

This approach ensures robust estimates of the ε_p values and their relevant CIs, enabling a statistically rigorous comparison of error rates among publishers. For instance, based on Table 4, it can be observed that – with respect to Scopus – the $\hat{\varepsilon}_p$ value for documents published by "Springer Nature" is significantly higher than those for "Elsevier", "MDPI" and "IEEE". The difference is further evidenced by the non-overlapping 95% CIs around the same $\hat{\varepsilon}_p$ values (see graphical representation via the boxplot in Fig. 2). Such non-overlap suggests that the "true" error rates for these publishers are statistically distinct, implying systematic differences in how documents from different publishers are classified.

Table 4 *– Summary of the publisher-specific analysis for the Scopus database (on the left) and WoS database (on the right). For each p-th publisher (listed in the table rows), the following values are reported: the total number of analyzed documents (d_p), the number of observed document-type classification errors (e_p), the corresponding (estimated) error rate ($\hat{\varepsilon}_p = \frac{e_p}{d_p}$), the 95% CI around $\hat{\varepsilon}_p$ (see Eq. 4), and the deviation (dev_p) associated with the χ^2 test, which accounts for differences among publishers (see Eqs. 7 and 8). The bottom row reports the overall values for the entire document corpus.*

		Scopus					WoS				
				95% CI					95% CI		
Publisher	d_p	e_p	$\hat{\varepsilon}_p$	Lower limit	Upper limit	dev_p	e_p	$\hat{\varepsilon}_p$	Lower limit	Upper limit	dev_p
(A) Elsevier	2,735	37	0.014	0.009	0.018	7.195	32	0.012	0.008	0.016	0.231
(B) MDPI	1,812	22	0.012	0.007	0.017	6.714	21	0.012	0.007	0.017	0.188
(C) IEEE	1,648	29	0.018	0.011	0.024	0.886	15	0.009	0.005	0.014	1.710
(D) Springer Nature	1,290	59	0.046	0.034	0.057	37.817	19	0.015	0.008	0.021	0.401
(E) John Wiley & Sons, Inc.	515	18	0.035	0.019	0.051	4.817	5	0.010	0.001	0.018	0.371
(F) Taylor & Francis Group	321	8	0.025	0.008	0.042	0.241	8	0.025	0.008	0.042	3.741
(G) IOP Publishing Ltd	308	11	0.036	0.015	0.056	3.203	8	0.026	0.008	0.044	4.236
Overall	**8,629**	**184**	**0.021**			**60.873**	**108**	**0.013**			**10.879**

The chi-squared (χ^2) test confirms these considerations, evaluating whether the observed error rates differ significantly across publishers. For each *p*-th publisher, the deviation (dev_p, representing the standardized residuals of the χ^2 test) is calculated as follows [10]:

$$dev_p = \frac{(O_p - E_p)^2}{E_p}, \tag{5}$$

where $O_p = e_p = \varepsilon_p \cdot d_p$ is the *observed* number of errors and $E_p = \varepsilon \cdot d_p$ is the *expected* number of errors under the null hypothesis (H_0) of equal error rates across publishers.
The deviation relative to the *p*-th publisher therefore becomes:

$$dev_p = \frac{(\varepsilon_p \cdot d_p - \varepsilon \cdot d_p)^2}{\varepsilon \cdot d_p} = \frac{(\varepsilon_p - \varepsilon)^2}{\varepsilon} \cdot d_p. \tag{6}$$

The test statistic χ^2 is then computed as:

$$\chi^2 = \sum_{\forall p} dev_p = \sum_{\forall p} \frac{(\varepsilon_p - \varepsilon)^2}{\varepsilon} \cdot d_p, \tag{7}$$

which is compared against the critical value $\chi^2_{\alpha,\, DoF}$ for the chosen significance level (e.g., α = 5%)

Materials Research Proceedings 57 (2025) 80-89 https://doi.org/10.21741/9781644903735-10

and *degrees of freedom*, determined as:

$$DoF = (rows - 1) \cdot (colums - 1), \tag{8}$$

as a function of the number of rows and columns of the contingency table containing the data of interest. In this case, *rows* = 2 (distinguishing between correctly classified and misclassified documents), while *columns* is equal to the number of publishers, determining $DoF = (2 - 1)\cdot(7 - 1) = 6$. The comparison between the calculated χ^2 value and the critical one leads to two possible outcomes:

$$\begin{gathered}\chi^2 > \chi^2_{\alpha,DoF} \text{ (or } p < \alpha\text{): Reject the null hypothesis (H}_0\text{);} \\ \chi^2 \leq \chi^2_{\alpha,DoF} \text{ (or } p \geq \alpha\text{): Do not reject H}_0\text{,}\end{gathered} \tag{9}$$

where H_0 is that there are no significant differences between publishers in terms of DT-classification error rates. In the specific case under analysis, the result for Scopus is:

$$\chi^2 = 60.873 > \chi^2_{\alpha,\,DoF} = 12.592 \;\; \left(p = 3 \times 10^{-11} < \alpha = 0.05\right), \tag{10}$$

which leads to rejecting H_0. From the dev_p values (cf. Table 4 on the left) it becomes evident that certain publishers deviate more significantly than others. For instance, “Springer Nature”, “IOP Publishing Ltd” and “John Wiley & Sons, Inc.” have significantly higher ε_p values, while “Elsevier”, “MDPI” and “IEEE” exhibit significantly lower values compared to the rest.

In WoS (cf. Table 4 on the right), the χ^2 test does not indicate significant differences among publishers:

$$\chi^2 = 10.879 \leq \chi^2_{\alpha,\,DoF} = 12.592 \;\; (p = 0.09 \geq \alpha = 0.05), \tag{11}$$

which leads to failing to reject H_0, meaning that there is no strong statistical evidence of significant differences in DT-classification error rates among publishers. This result appears to be further supported by the diagram in Fig. 2 on the right, which shows that the CIs around the ε_p values for different publishers in WoS tend to overlap, albeit slightly.

Table 5 includes an additional test on pairwise differences in error rates. Specifically, for each pair of publishers (e.g., considering two generic publishers *A* and *B*), the null hypothesis (H_0) assumes that the two error rates (ε_p) are equal, while the alternative hypothesis (H_1) assumes they differ. The test statistic *z* is defined as follows:

$$z = \frac{\varepsilon_A - \varepsilon_B}{\sqrt{\varepsilon_{pool}\cdot(1-\varepsilon_{pool})\cdot\left(\frac{1}{d_A}+\frac{1}{d_B}\right)}}, \tag{12}$$

where:

ε_A and ε_B are the observed error rates for the two publishers, calculated as: $\varepsilon_A = \frac{e_A}{d_A}$, $\varepsilon_B = \frac{e_B}{d_B}$;

d_A and d_B are the numbers of documents analyzed for each publisher;

ε_{pool} is the pooled error rate, computed as: $\varepsilon_{pool} = \frac{e_A + e_B}{d_A + d_B}$.

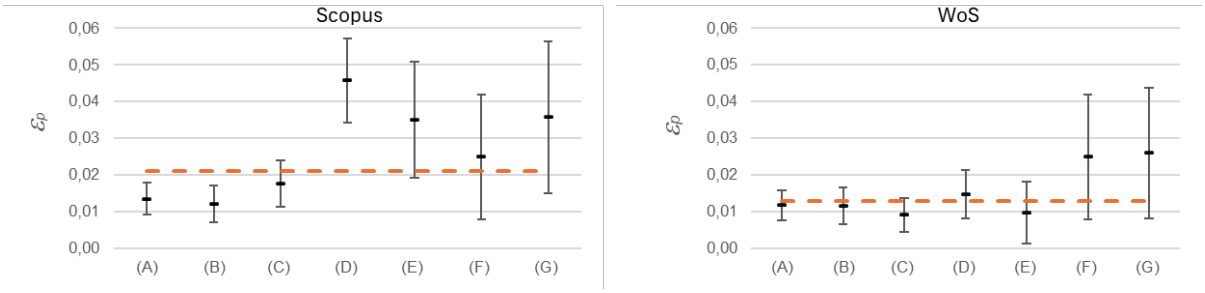


Figure 2 *– ε_p values and corresponding 95% CIs for Scopus (on the left) and WoS (on the right). The numerical values are provided in Table 4. The horizontal dashed line represents the overall error rate (ε) of the referenced database. On the x-axis the related publisher listed in descending order based on the total number of documents included in the analysis (cf. Fig. 1 and Table 4).*

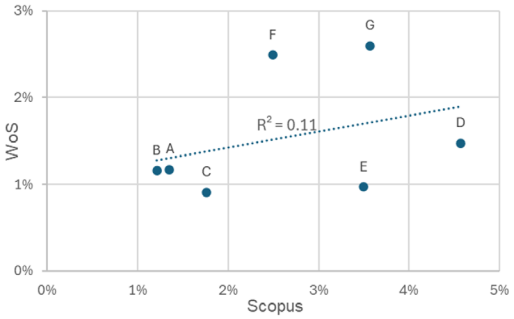
Table 5 *– Contingency table of z-values for pairwise comparisons of error rates. Values where* $|z| > 1.96$ *are highlighted in bold.*

		(A) Elsevier	(B) MDPI	(C) IEEE	(D) Springer Nature	(E) John Wiley & Sons, Inc.	(F) Taylor & Francis Group	(G) IOP Publishing Ltd
Scopus	(A) Elsevier	0.00	0.40	-1.07	**-6.25**	**-3.46**	-1.60	**-2.96**
	(B) MDPI	-0.40	0.00	-1.33	**-5.78**	**-3.51**	-1.79	**-3.09**
	(C) IEEE	1.07	1.33	0.00	**-4.44**	**-2.36**	-0.88	**-2.06**
	(D) Springer Nature	**6.25**	**5.78**	**4.44**	0.00	1.02	1.67	0.77
	(E) John Wiley & Sons, Inc.	**3.46**	**3.51**	**2.36**	-1.02	0.00	0.81	-0.06
	(F) Taylor & Francis Group	1.60	1.79	0.88	-1.67	-0.81	0.00	-0.79
	(G) IOP Publishing Ltd	**2.96**	**3.09**	**2.06**	-0.77	0.06	0.79	0.00
WoS	(A) Elsevier	0.00	0.03	0.81	-0.80	0.39	**-1.97**	**-2.09**
	(B) MDPI	-0.03	0.00	0.72	-0.76	0.36	-1.90	**-2.01**
	(C) IEEE	-0.81	-0.72	0.00	-1.42	-0.13	**-2.41**	**-2.52**
	(D) Springer Nature	0.80	0.76	1.42	0.00	0.84	-1.27	-1.38
	(E) John Wiley & Sons, Inc.	-0.39	-0.36	0.13	-0.84	0.00	-1.73	-1.81
	(F) Taylor & Francis Group	**1.97**	1.90	**2.41**	1.27	1.73	0.00	-0.08
	(G) IOP Publishing Ltd	**2.09**	**2.01**	**2.52**	1.38	1.81	0.08	0.00

In practice, after computing the *z*-statistic (Eq. 12), it is compared with the critical value $z_{1-\frac{\alpha}{2}} \approx 1.96$ for $a = 5\%$. If $|z| > 1.96$, the null hypothesis is rejected, indicating that the two error rates (ε_A and ε_B) are statistically different [10]. The procedure is then repeated for all $\binom{7}{2} = 21$ possible pairwise comparisons. With regard to Scopus, three pairwise comparisons (bolded) show statistically significant differences, confirming the findings of the previous statistical tests. The pairwise comparisons revealing the largest differences involve publishers "Springer Nature", "John Wiley & Sons, Inc." and "IOP Publishing Ltd". With regard to WoS, it is observed that "Taylor and Francis Group" and "IOP Publishing Ltd" exhibit statistically significant differences compared with the other publishers, as they tend to have higher ε_p values (cf. boxplots in Fig. 2). The results from each database's perspective can be summarized into the map in Fig. 3, which plots the ε_p values for Scopus on the *x*-axis and those for WoS on the *y*-axis, for each of the 7 (macro-)publishers. Notably, there appears to be no significant correlation between the publishers that are most "critical" for one database (in terms of ε_p values) and those that are most critical for the other, as indicated by the relatively low $R^2 = 0.11$. This finding suggests that the two databases may differ in their familiarity with handling DTs for specific publishers. The causes of this apparent independence will be investigated in future studies, including: (a) the two databases may adopt different DT classification rules, which could align closely with the editorial standards of specific publishers, leading to varying error patterns and (b) some publishers may have a closer relationship with a given database (e.g., Elsevier and the Scopus database are part of the same organization

Materials Research Proceedings 57 (2025) 80-89 https://doi.org/10.21741/9781644903735-10

(RELX Group), which could reduce the likelihood of classification errors or facilitate their prompt correction when they occur) [11].



Publisher		Scopus	WoS
(A)	Elsevier	1.4%	1.2%
(B)	MDPI	1.2%	1.2%
(C)	IEEE	1.8%	0.9%
(D)	Springer Nature	4.6%	1.5%
(E)	John Wiley & Sons, Inc.	3.5%	1.0%
(F)	Taylor & Francis Group	2.5%	2.5%
(G)	IOP Publishing Ltd	3.6%	2.6%

Figure 3 – *Map comparing the* $\boldsymbol{\varepsilon_p}$ *results for the two databases: Scopus (x-axis) and WoS (y-axis).*

4. Concluding remarks

This research yielded several findings of interest for academic researchers in the field of engineering/manufacturing. Firstly, while DT-classification errors can be considered relatively rare events, their occurrence is nonetheless significant in both Scopus (error rate ~2.1%) and WoS (error rate ~1.3%) (cf. ***RQ1***). In addition to influencing various types of academic assessments—such as evaluations of individual researchers for career advancement, rankings of groups of researchers for funding calls, etc.—these errors could lead to more specific effects in the engineering/manufacturing sector. For instance, they could cause the distortion of metrics related to technology transfer and industrial innovation of research groups/institutions.

A deeper analysis from the publishers' perspective revealed statistically significant differences in their propensity to incur DT-classification errors. Specifically, Scopus appears to struggle more with publishers "Springer Nature", "IOP Publishing Ltd" and "John Wiley & Sons, Inc.", with relevant ε_p values of ~4.6%, ~3.6% and ~3.5% respectively, whereas WoS shows higher error rates with publishers "IOP Publishing Ltd" and "Taylor & Francis Group", with relevant ε_p values of ~2.6% and ~2.5% respectively (cf. ***RQ2***). The reasons behind these discrepancies remain to be thoroughly investigated, but they could include various factors (cf. ***RQ3***). Possible causes include differences in editorial conventions, alignment with database classification rules, and the diversity of DTs within publishers' portfolios (i.e., larger publishers span a broader range of disciplines—even within the same field—and manage extensive portfolios of journals and books; this increased diversity of DTs may lead to a higher likelihood of DT-classification errors). Mergers and acquisitions (M&A), where smaller publishers are absorbed by larger groups, may also influence DT-classification consistency, as editorial standards can either remain unchanged or be adapted to align with the acquiring publisher's guidelines.

From a practical standpoint, this work aims to raise awareness of the possibility of DT-classification errors and the fact that, depending on the database in use, documents from certain publishers may be at a higher risk of misclassification. In general, although Scopus and WoS do not seem to routinely conduct periodic reviews of the accuracy of indexed data, such as by cross-referencing results with those of competing databases, they do appear to be relatively responsive in addressing reported errors through designated channels [12].

This study has some limitations, primarily its focus on engineering/manufacturing, which may restrict the generalizability of the findings to other disciplines [13,14]. Additionally, some documents were excluded, such as those without DOI codes or those assigned multiple DT-classifications in WoS. Future research will seek to expand the sample to a broader range of

scientific fields and further investigate the underlying causes of publisher-specific classification discrepancies.

Acknowledgements
This study was carried out within the MICS (Made in Italy – Circular and Sustainable) Extended Partnership and was partially funded by the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1551.11-10-2022, PE00000004). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

References

[1] Harzing, A. W. (2013). Document categories in the ISI Web of Knowledge: Misunderstanding the social sciences?. Scientometrics, 94(1), 23-34. https://doi.org/10.1007/s11192-012-0738-1

[2] Donner, P. (2017). Document type assignment accuracy in the journal citation index data of Web of Science. Scientometrics, 113(1), 219-236. https://doi.org/10.1007/s11192-017-2483-y

[3] Franceschini, F., Maisano, D., Mastrogiacomo, L. (2015). Influence of omitted citations on the bibliometric statistics of the major Manufacturing journals. Scientometrics, 103(3), 1083-1122. https://doi.org/10.1007/s11192-015-1583-9

[4] García-Pérez M.A. (2010) Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in psychology. Journal of the American society for information science and technology, 61/10: 2070-85. https://doi.org/10.1002/asi.21372

[5] Sigogneau, A. (2000). An analysis of document types published in journals related to physics: Proceeding papers recorded in the Science Citation Index database. Scientometrics, 47(3), 589-604. https://doi.org/10.1023/A:1005628218890

[6] Mokhnacheva, Y. V. (2023). Document Types Indexed in WoS and Scopus: Similarities, Differences, and Their Significance in the Analysis of Publication Activity. Scientific and Technical Information Processing, 50(1), 40-46. https://doi.org/10.3103/S0147688223010033

[7] Maisano, D.A., Mastrogiacomo, L., Ferrara L., and Franceschini F. (2025) A large-scale semi-automated approach for assessing document-type classification errors in bibliometric databases. Scientometrics, 130, 1901-1938. https://doi.org/10.1007/s11192-025-05244-y

[8] Franceschini, F., Maisano, D. (2014). Sub-field normalization of the IEEE scientific journals based on their connection with Technical Societies. Journal of Informetrics, 8(3), 508-533. https://doi.org/10.1016/j.joi.2014.04.005

[9] Haupka, N., Culbert, J. H., Schniedermann, A., Jahn, N., Mayr, P. (2024). Analysis of the Publication and Document Types in OpenAlex, Web of Science, Scopus, PubMed and Semantic Scholar. arXiv preprint arXiv:2406.15154.

[10] Ross, S.M. (2017). Introductory statistics. Academic Press (Elsevier), London. https://doi.org/10.1016/B978-0-12-804317-2.00031-X

[11] Franceschini, F., Maisano, D., Mastrogiacomo, L. (2016). Empirical analysis and classification of database errors in Scopus and Web of Science. Journal of informetrics, 10(4), 933-953. https://doi.org/10.1016/j.joi.2016.07.003

[12] Franceschini, F., Maisano, D., & Mastrogiacomo, L. (2016). Do Scopus and WoS correct "old" omitted citations? Scientometrics, 107, 321-335. https://doi.org/10.1007/s11192-016-1867-8

[13] Franceschini, F., & Maisano, D. (2011). Bibliometric positioning of scientific manufacturing journals: A comparative analysis. Scientometrics, 86(2), 463-485. https://doi.org/10.1007/s11192-010-0301-x

[14] Franceschini, F., Maisano, D., Turina, E. (2012). European research in the field of production technology and manufacturing systems: an exploratory analysis through publications and patents. The International Journal of Advanced Manufacturing Technology, 62(1-4), 329-350. https://doi.org/10.1007/s00170-011-3791-7