

**Editor**

Valéria dos Santos Gouveia Martins

**Support**

Fundação de Amparo à Pesquisa do Estado do Maranhão (Process No APP-09405/22) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) - Finance Code 001.

**Conflict of interest**

The authors declare that there are no conflicts of interest.

**Data Availability**

The research data are available from the corresponding author upon reasonable request.

**Received**

February 8, 2025

**Final version**





April 2, 2025

**Approved**

April 20, 2025

# Career-long data outperforms recent data in predicting the publication success of physicists

## *Dados de toda a carreira superam dados recentes na predição do sucesso de publicações de físicos*

Inez Cavalcanti Dantas<sup>1</sup> , Antonio de Abreu Batista-Júnior<sup>1</sup> , Luciano Reis Coutinho<sup>1</sup> , Jesús Pascual Mena-Chalco<sup>2</sup> 

<sup>1</sup> Universidade Federal do Maranhão, Centro de Ciências Exatas e Tecnológicas, Programa de Pós-Graduação em Ciência da Computação. São Luís, MA, Brasil. Correspondence to: A. A. Batista-Junior. E-mail: <antonio.batista@ufma.br>.

<sup>2</sup> Universidade Federal do ABC, Centro de Matemática, Computação e Cognição, Programa de Pós-Graduação em Ciência da Computação. Santo André, SP, Brasil.

**How to cite this article:** Dantas, I. C. *et al.* Career-long data outperforms recent data in predicting the publication success of physicists. *Transinformação*, v. 37, e2515056, 2025. <https://doi.org/10.1590/2318-0889202537e2515056>

### Abstract

Predicting the future success of researchers is an important topic that is attracting increasing attention in the research community. However, there are no comprehensive studies that show which predictive bibliometric indices best predict the short-term success of researchers. It does not matter whether these are metrics based on the researcher's most recent publications or metrics calculated based on all publications throughout the researcher's career. In this study, we are interested in how the two time windows used as the basis for calculating predictive bibliometric indices affect the empirical results of a classifier in predicting the success of researchers. Using the American Physical Society dataset, we compare the performance of the classifiers using 10-fold cross-validation to determine the most accurate classifier in both scenarios described above. The results of this study disprove our initial hypothesis that it is better to train with bibliometric indices computed from recent publications. This result suggests that there is a need to assess scientists more comprehensively, as focusing on recent publications and not taking into account results from earlier stages of their career may lead to poorer results in predicting scientists' publication success.

**Keywords:** Machine learning. Multilayer perceptron. Neural networks. Physicists. Success prediction.

### Resumo

*Prever o sucesso futuro dos pesquisadores é um tópico crítico que está atraindo cada vez mais atenção na comunidade de pesquisa. No entanto, não há estudos completos que demonstrem quais índices bibliométricos preditivos melhor predizem o sucesso de curto prazo de pesquisadores, independentemente de serem métricas baseadas nas publicações mais recentes do pesquisador ou métricas calculadas com base em todas as publicações ao longo da carreira do pesquisador. Neste estudo, estamos interessados em como as duas janelas de tempo usadas como base para calcular índices bibliométricos preditivos afetam os resultados empíricos de um classificador na*

*previsão do sucesso dos pesquisadores. Usando o conjunto de dados da American Physical Society, comparamos o desempenho dos classificadores. Usamos validação cruzada de 10 folds para determinar o classificador mais preciso em ambos os cenários descritos acima. Os resultados deste estudo refutam nossa hipótese inicial de que é melhor treinar com índices bibliométricos calculados a partir de publicações recentes. Esse resultado sugere que há uma necessidade de avaliar os cientistas de forma mais abrangente, pois focar-se em publicações recentes e não levar em consideração resultados de um estágio anterior da carreira pode levar a resultados inferiores na previsão do sucesso de publicações de cientistas.*

**Palavras-chave:** *Aprendizado de máquina. Perceptron multicamadas. Redes neurais. Físicos. Predição de sucesso.*

## Introduction

With the rapid development of scientific research and the equally rapid growth in the number of scientists, a huge number of publications are published every year. In this situation, the question of how to effectively predict the future influence of scientific papers and researchers has become an important research problem. The topic attracts the attention of researchers in various fields and is of great importance for improving research efficiency and supporting decision-making and scientific evaluation (Xia; Li; Li, 2023).

The study of temporal aspects of science has always been of interest to researchers and has attracted much attention in the past. Longitudinal studies of scientific data dominate most journal volumes in the respective scientific field (Manolopoulos *et al.*, 2021). Such studies attempt to uncover hidden patterns or useful insights for scientists, articles, journals and other stakeholders. However, are even more interesting approaches that attempt to project the dynamics of research impact into the future, in order to estimate as accurately as possible the future state and/or evolution of the data in question.

This is due to the fact that there are many real-world applications that can greatly benefit from accurate prediction of scientific dynamics. Take, for example, a system that recommends articles to scientists according to their research interests. As the number of published articles is rapidly increasing, a large number of suggestions are made on almost any topic. The recommendation system could prioritize the list of suggested articles based on estimates of the expected impact of the articles and give higher priority to those with a higher estimated impact. Similar benefits naturally arise for other use cases such as expert search, collaboration recommendations, etc.

To perform these applications, many attempts have been made in the past to predict the future performance (or influence) of entities, such as research articles (Bai; Zhang; Lee, 2019; Hirako; Sasano; Takeda, 2023) and scholars (Acuna; Allesina; Kording, 2012; Ayaz; Masood; Islam, 2018). Recent findings (Zhao; Feng, 2022) show that neural networks perform this task better in practice than other models. However, the “black box” nature of these networks has limited their use.

A neglected area in the field of predicting scientific impact is the question of whether bibliometric indices based only on recent data provide better predictions than using long-term data. To address this gap, in this study we aim to find out what is the optimal time window for predicting the short-term success of researchers. We hypothesize that classifiers trained with the first time window will perform better on this task than those trained with the second window. In this paper, the success (Frietsch; Gruber; Bornmann, 2025) of physicists is measured by the number of new future publications, where each has the same number of citations, in a short period of time.

To confirm our hypothesis, we compared the performance of Artificial Neural Networks (ANNs) as classifiers. We used the American Physical Society (APS) dataset and to determine the most accurate ANN in the two time windows, we used k-fold cross-validation (k=10). The results of

this study suggest that our hypothesis that it is better to train with indices calculated from last year's data – is incorrect. This result contradicts our practical intuition of considering recent performance as a proxy for the near future performance, rather than considering the entire career as a more reliable measure.

## Related Work

The prediction of scientific impact has long been a central challenge in scientometrics, with early studies introducing citation-based measures to assess researchers' influence. One of the most notable contributions in this field was Hirsch's (2005) development of the h-index, a metric that integrates both publication count and citation impact, as cited in Abramo, D'Angelo and Felici (2019). Since then, numerous alternative indices, such as the g-index and m-index, have been proposed to address the limitations of the h-index, incorporating factors like publication longevity and co-authorship weighting (Koltun; Hafner, 2021).

Over the years, various approaches have been explored to enhance bibliometric predictions. Traditional models have predominantly relied on citation counts and author-based attributes, as demonstrated by studies leveraging historical citation data to forecast future impact (Acuna; Allesina; Kording, 2012; Abramo; D'Angelo; Felici, 2019). In parallel, network-based methodologies have gained traction, utilizing co-authorship and citation networks to infer influence scores and predict scholarly success (Sinatra *et al.*, 2016). More recently, machine learning techniques have been applied in this domain, integrating features such as early citation growth, author reputation, and journal impact factors to improve predictive accuracy (Xia; Li; Li, 2023).

Advancements in impact prediction have also been driven by deep learning models and altmetric indicators. Research indicates that combining early citation patterns with machine learning algorithms can significantly enhance long-term impact predictions (Akella *et al.*, 2021). Furthermore, altmetrics – including social media mentions, downloads, and media coverage – have been explored as complementary indicators of scholarly influence, reflecting a broader shift from traditional citation-based metrics to more diverse and holistic evaluation frameworks (Marques-Cruz *et al.*, 2024).

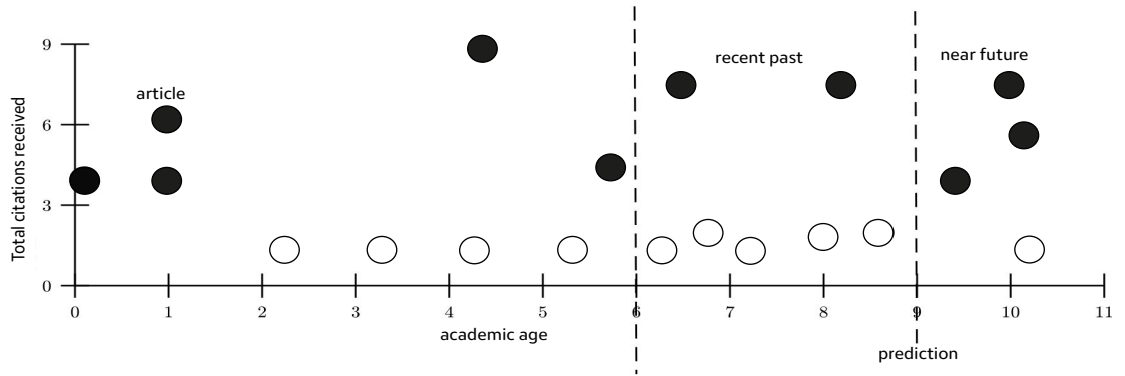
Despite these innovations, several critical challenges remain. A key gap in the literature is the lack of consensus on the optimal time window for calculating predictive bibliometric indices. While many studies assume that recent publications are the strongest predictors of future success, empirical evidence suggests otherwise (Teplitskiy *et al.*, 2022). Additionally, although machine learning models have enhanced predictive accuracy, their "black box" nature limits interpretability and trust in their predictions (Koltun; Hafner, 2021).

This study seeks to address these challenges by systematically evaluating different time windows for bibliometric predictions and assessing their empirical effectiveness in forecasting scientific success. By leveraging a comprehensive dataset and employing rigorous evaluation methods, this research provides new insights into the temporal dynamics of scholarly impact assessment.

## Methodological Procedures

The success of physicists in the near future is measured by the success of their scientific publications after the year marked as the prediction year (Figure 1). The success of physicists in the past is measured by the success of their scientific publications prior to the year marked as the

prediction year. We use these criteria to create a dataset of examples of successful and unsuccessful physicists and train an artificial neural network that can predict new, never-before-seen examples. We define this as a binary classification problem, where a successful physicist is defined as someone who has published at least a number of publications, each with at least the same number of citations, in the years following the prediction year, and an unsuccessful physicist is the opposite.



**Figure 1** – The publication list of a successful scientist, because he has at least three articles with three citations each in the near future.

Source: Prepared by the authors (2025).

We use the following bibliometric indices of the scientist as predictor features:  $H$ , the H-index of the scientist  $u$  from the specific year  $y$  of the career to the year of the prediction;  $co$ , the number of different co-authors of scientist  $u$  from year  $y$  of career to year of prediction;  $c$ , the total number of citations of scientist  $u$  from year  $y$  of career to year of prediction;  $I_i^u$ , the total number of articles by scientist  $u$  with  $i$  citations each from year  $y$  of career to year of prediction; and  $v$ , the total number of unique journals in which scientist  $u$  has published from year  $y$  of his career to the year of prediction.

We compare the effects of a short time window for the collection of publications, which serves as the basis for the calculation of the indices, with a long time window representing all publications of the scientist, on the predictive power of the predictors  $H$ ,  $co$  and  $c$  on the performance of the classifier.

We used multiple feedforward neural networks with two hidden layers as classifiers, where  $k$  and  $L$  are the number of neurons in the first and second hidden layers. The output of the networks uses a logistic function and the output of the hidden layers uses Rectified Linear Unit (RELU).

## Feature engineering

The bibliometric indices discussed in the previous section were engineered as follows: let  $A_{n \times p}$  be a matrix in which an entry  $a_{ij}$  indicates the number of citations that article  $i$  received in year  $j$ .  $n$  is the number of articles belonging to the dataset, and  $p$  is the total number of years after the year of publication.

The total number of citations of article  $i$  in the data set is  $c_i = \sum_{j=0}^p a_{ij}$ , where  $p$  is the maximum number of years. The total number of citations of a scientist  $u$  in the data set is  $c_u = \sum_{i \in L_u} c_i$ , where  $L_u$  is the list of publications of  $u$ .

The  $H_u$ -index of a scientist  $u$  is calculated based on the basis of his publication list  $L_u$ . The formula is based on the number of articles ( $H$ ) that have been cited a minimum number of times, and how often, compared to those that have not been cited (or have not been cited as often).

For example, a researcher has an H-index of 6 because they have 6 publications that have been cited at least 6 times. The remaining articles or those that have not yet been cited 6 times are left aside. The number of co-authors of scientist  $u$  was calculated using the co-authorship matrix  $C_{n \times m}$ , where an entry  $c_{ui}=1$  indicates that author  $u$  wrote paper  $i$ , and  $c_{ui}=0$  otherwise.

## Performance evaluation metrics

The performance of the models was measured using the Matthews correlation coefficient (Thai-Nghe; Gantner; Schmidt-Thieme, 2010) and the f1 score (Hand; Christen; Kirielle, 2021) in a k-fold cross-validation. We chose these measures because they are the most accurate when evaluating classifiers trained with a significantly greater number of examples of one class than the other. Only 20 percent of our examples come from short-term successful scientists.

We addressed class imbalance with resampling techniques and cost-sensitive learning. The former focuses on manipulations at the data level and the latter targets on the internal changes in the classifier (Thai-nghe; Gantner; Schmidt-Thieme, 2010).

To test binary classifiers on imbalanced datasets, accuracy metrics and f1-scores calculated from confusion matrices are among the most commonly used metrics in binary classification tasks. However, these statistical measures can lead to dangerously inflated results, especially for imbalanced datasets (Chicco; Jurman, 2020). For this reason, we use the Matthews Correlation Coefficient (MCC), a more reliable statistical measure that only results in a high score if the prediction performs well in all four categories of the confusion matrix (True Positives (TP), False Negatives (FN), True Negatives (TN), and False Positives (FP)). The MCC is the only performance metric for binary classification that yields a high score only when the binary classifier is able to correctly predict most instances of both the majority and the minority classes.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP) \times (TP+FN) \times (TN+FP) \times (TN+FN)}}$$

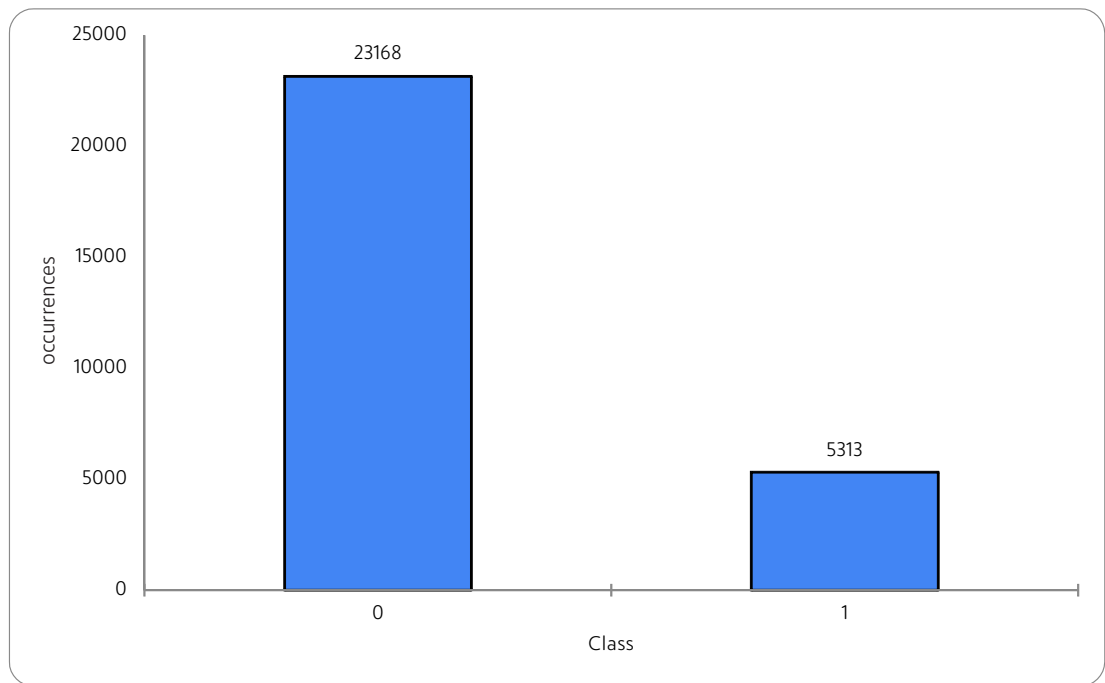
However, we also use the f1 score as a second, but not the main, performance measure.

$$f1 = \frac{(2 \times TP)}{FN + FP + 2 \times TP}$$

## Dataset

We used the dataset managed by the APS (<https://journals.aps.org/datasets>) as the data source for our analyses and used the author names from the article metadata to identify the authors. We defined the academic age of authors based on the number of years between their first and last publication. We selected authors who started publishing after 1985 and had an academic age of at least 8 years. We used their publications up to three years before their maximum age to calculate the value of the predictors and their remaining publications as future publications. We used all citations for a given article within the dataset and therefore did not define a time window for the accumulation of citations.

The dataset is unbalanced. Approximately 20 percent (Figure 2) are examples of successful physicists (class 1) and the rest are examples of unsuccessful physicists.



**Figure 2** – The uneven distribution of classes in the dataset.  
 Source: Prepared by the authors (2025).

## Class imbalance

Class imbalance refers to the uneven distribution of classes, i.e. there are too many examples in one class and too few in others. This affects the performance of standard learning algorithms that require a balanced distribution of classes. There are different approaches to address the class imbalance problem: at the data level (changing the distribution of the training set to restore balance by adding or removing instances from the training dataset), at the algorithm level (changing the objective function of the classifier to increase the importance of the minority class), and with hybrid methods (combining algorithm-level methods with data-level approaches). We addressed class imbalance at the data and algorithm levels. At the data level, we use under-sampling of the majority class and removed instances from the training dataset to restore the balance. At the algorithm level, we use a cost-sensitive learning approach that specifies a loss function that considers a higher weight for positive (minority) examples.

## Results and Discussion

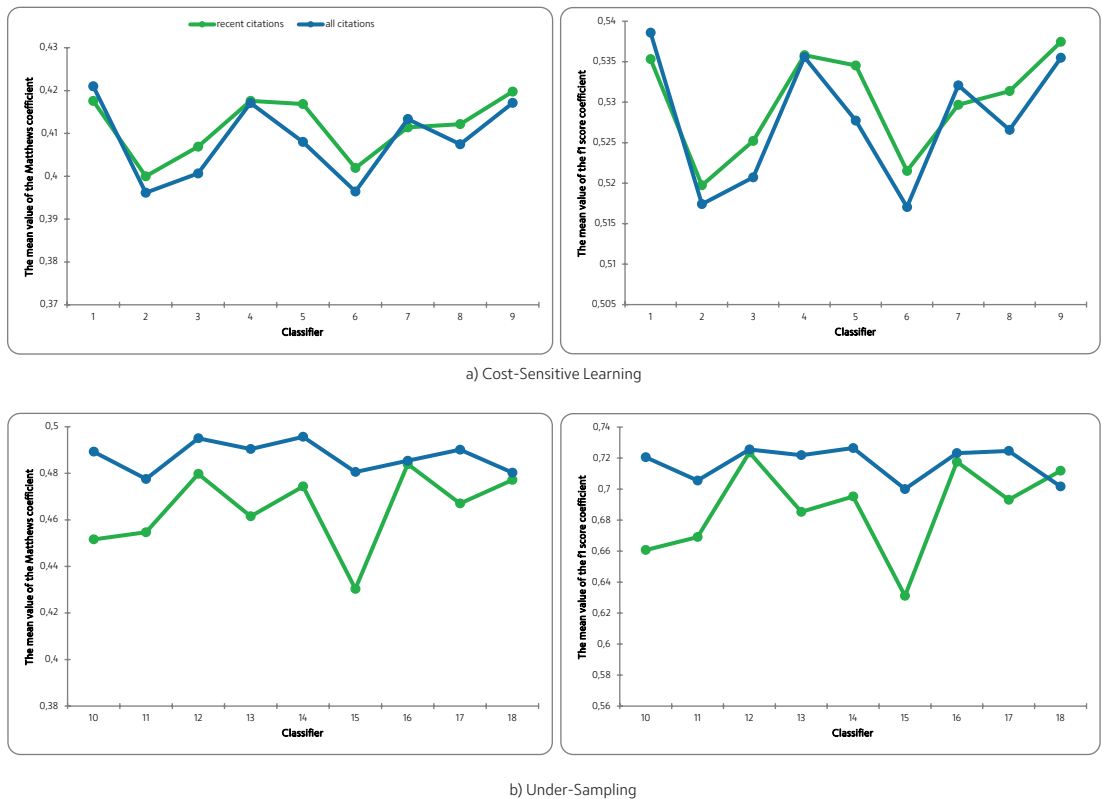
To test our hypothesis that bibliometric indices calculated using fresh data from the scientist's most recent publications are better predictors of successful physicists than those calculated using publications from their entire career, we evaluated the performance of predictive classifiers (Table 1) for predicting the total number of successful articles in the short term on the APS dataset. We compared the performance of these classifiers with the total citation index calculated based on the scientist's publications in the last three years and with this index calculated based on all of the scientist's publications. The other predictors (mixed) are identical in both cases to assess only the influence of this index. We used the mean of a k-fold cross-validation (k=10) to determine the most accurate classifier and indirectly the effects of these two decisions.

**Table 1** – Hyperparameter values.

Class Imbalance Techniques	NET	Number of neurons		Weight $w$
		Layer $j$	Layer $k$	
Cost-Sensitive Learning	1	25	25	4
	2	25	25	4.5
	3	25	25	5
	4	50	50	4
	5	50	50	4.5
	6	50	50	5
	7	100	100	4
	8	100	100	4.5
	9	100	100	5
Under-sampling the majority class	10	25	25	1*
	11	50	25	1
	12	100	25	1
	13	25	50	1
	14	50	50	1
	15	100	50	1
	16	25	100	1
	17	50	100	1
	18	100	100	1

Note: \*A weight of one means that both classes are equally important. NET: Neural Network.  
 Source: Prepared by the authors (2025).

As can be seen in Figure 3, the results of this experiment suggest that our hypothesis should be rejected. The average values of the performance in a k-fold cross-validation (k=10) between the classifiers trained on the dataset with the feature “Total citations received” for all publications



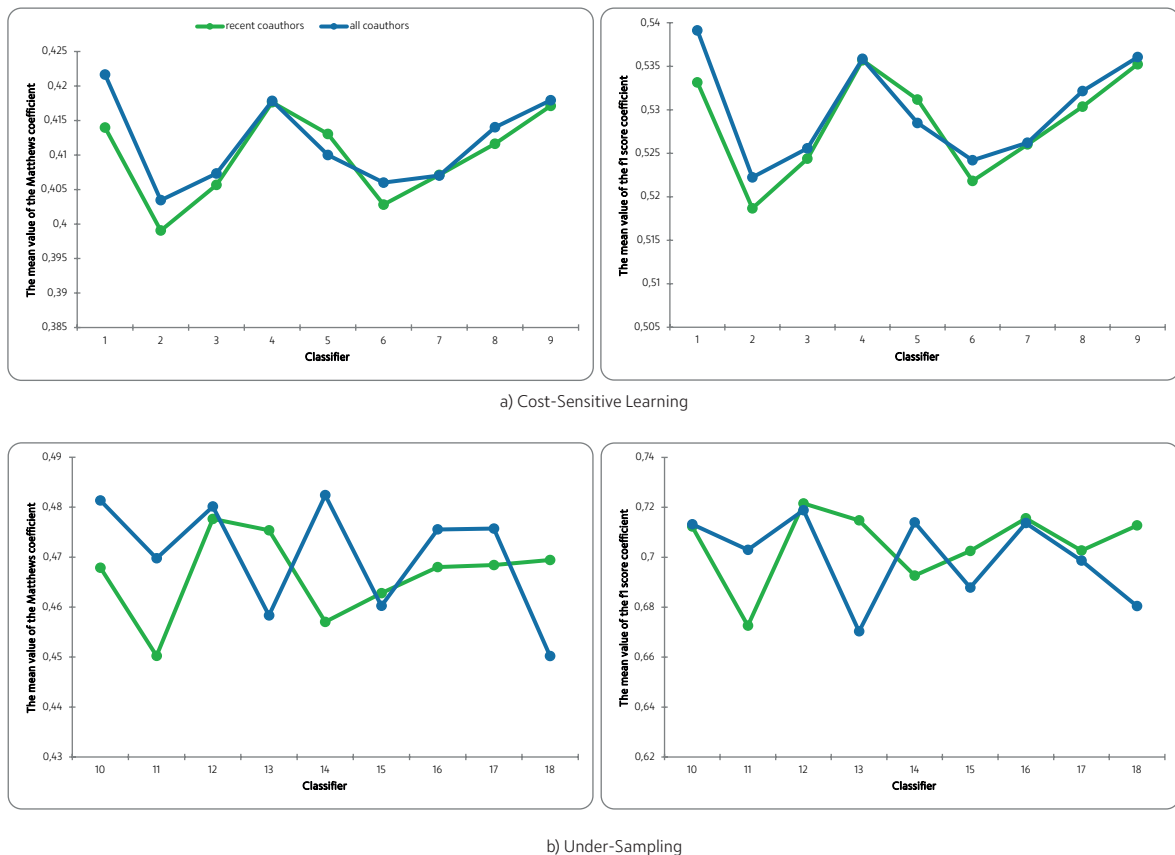
**Figure 3** – Comparison of the mean values between the performances of the individual classifiers in the 10-fold cross-validation, whereby the indicator, citations, was calculated with recent data and data from the entire career.

Source: Prepared by the authors (2025).

of the scientist are higher than those of the other classifiers trained with the feature “Total citations received” only for the most recent publications. Note that this result is independent of the method chosen to solve the class imbalance problem (under-sampling of the majority class and cost-sensitive learning approach). The total number of epochs and batch sizes for training in each fold were 300 epochs, each and the batch size was 100. A weight of 4 (Table 1) means that the loss behaves as if the dataset contains four times more positive examples (of the minority class).

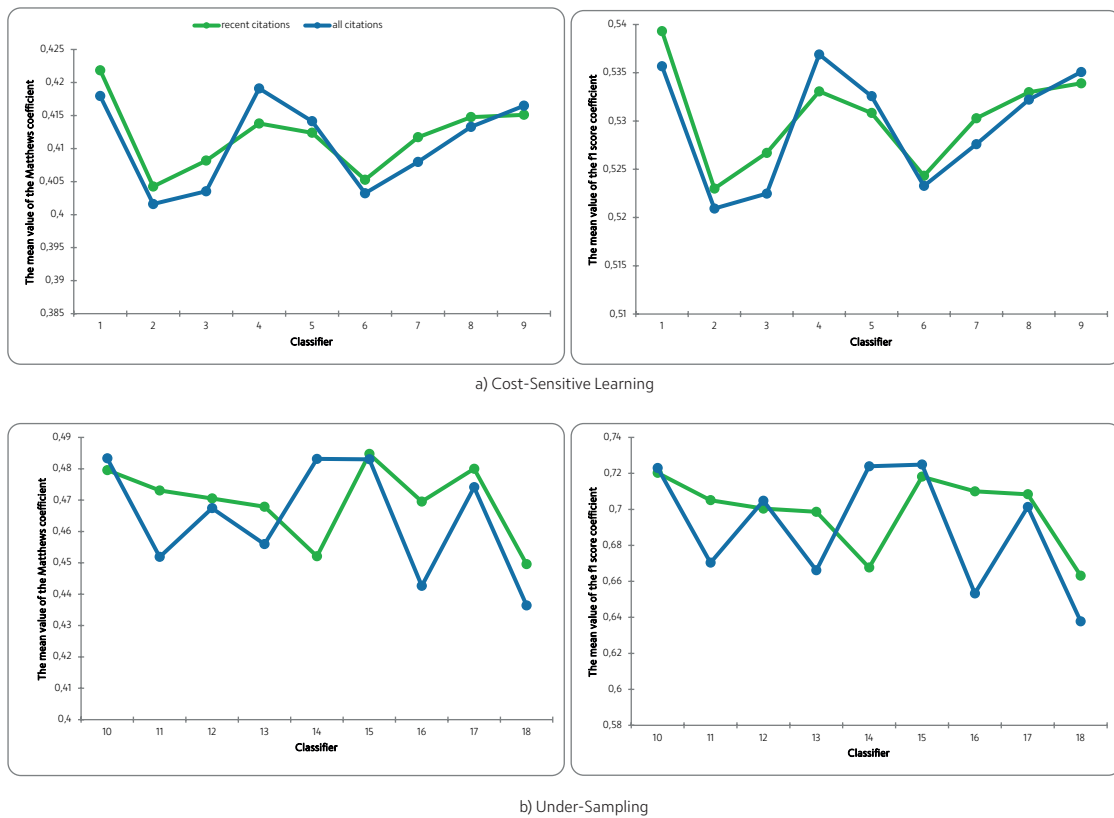
A higher Matthews correlation coefficient and a higher f1 value indicate a better performance of a classifier compared to a classifier with a lower value. Similarly, Figures 4 and 5 confirm a clear tendency to reject our hypothesis, regardless of the bibliometric index used as a predictor (co, H-index, citation). This suggests the opposite of what we expected, namely that it is better to use a short time window for the collection of publications than an entire window as the basis for calculating the predictive indicators. The results of this study argue in favor of a comprehensive time window that includes all articles in a career.

This work has shown that the size of the time window for the collection of publications used to calculate the indices is a critical factor for the accuracy of the predictors, but has been ignored so far. The main limitation of this study arises from the fact that the study focuses on a data set specific to physicists and the main observations of this study may not be generalizable to other fields.



**Figure 4** – Comparison of the mean values between the performances of the individual classifiers in the 10-fold cross-validation, whereby the indicator, the total number of coauthors (diferente), was calculated with recent data and data from the entire career.

Source: Prepared by the authors (2025).



**Figure 5** – Comparison of the mean values between the performances of the individual classifiers in the 10-fold cross-validation, whereby the indicator, H index, was calculated with recent data and data from the entire career.

Source: Prepared by the authors (2025).

Our experiments indirectly confirm previous findings (Sinatra *et al.*, 2016) that the process of predicting a high-impact scientific article among others in a career is not completely random, but can be partially predetermined. This conclusion is based on the fact that classifiers can successfully predict which scientists will publish high-impact articles in the short term. As Momeni, Mayr, and Dietze (2023) note, the evidence we found suggests that author-based metrics are good predictors of these researchers' own future metrics, but the time windows used to calculate them are crucial.

A major source of uncertainty lies in the nature of the indices used as predictors, which are based exclusively on citations of publications. Evaluation with indices that focus more on the metadata of publications is an important topic for future research.

We acknowledge that several additional factors merit consideration in the selection of the training corpus. These include the geographical affiliation of authors, the diversity of subfields within physics, the thematic scope of the relevant literature, the potential impact of researchers' administrative roles within their institutions, and other contextual variables that may influence research outcomes. Such limitations underscore the inherent complexity of assembling datasets that encompass not only publication metadata and citation networks, but also richer contextual information that is often difficult to obtain.

## Final Consideration

Our study demonstrates that computing predictive indices using a short publication collection time window in physics yields poorer classification performance than employing a longer

window. This outcome contradicts our initial hypothesis that a shorter time frame might produce more effective predictions. Furthermore, our results indicate that under-sampling is a more effective strategy for addressing class imbalances than the cost-sensitive Learning approach.

Despite the robustness of these findings, their generalizability across other disciplines remains uncertain. Additional experimental research is needed to validate these conclusions in various academic fields and to refine the estimation of the optimal time window for citation-based prediction.

Beyond methodological refinements, our research points to several promising avenues for future inquiry. Expanding the predictor set to include metadata-based indices (such as co-authorship networks, institutional affiliations, and funding sources) could enhance the precision of success predictions. Moreover, integrating deep learning techniques, particularly explainable AI models, may provide deeper insights into the complex factors influencing a researcher's future impact.

In a broader context, this study contributes to the ongoing discourse on predictive bibliometrics by underscoring the importance of comprehensive evaluation metrics. Identifying optimal conditions for forecasting academic success can facilitate more effective resource allocation, career planning, and institutional assessment. Ultimately, refining these predictive models will enhance transparency in scientific evaluation, mitigate biases in research assessment, and promote a more equitable academic landscape.

## References

- Acuna, D. E.; Allesina, S.; Kording, K. P. Predicting scientific success. *Nature*, v. 489, n. 7415, p. 201-202, 2012. Doi: <https://doi.org/10.1038/489201a>.
- Abramo, G.; D'Angelo, C. A.; Felici, G. Predicting publication long-term impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, v. 13, n. 1, p. 32-49, 2019. Doi: <https://doi.org/10.1016/j.joi.2018.11.003>.
- Akella, A. P. et al. Early indicators of scientific impact: predicting citations with altmetrics. *Journal of Informetrics*, v. 15, n. 2, p. 101128, 2021. Doi: <https://doi.org/10.1016/j.joi.2020.101128>.
- Ayaz, S.; Masood, N.; Islam, M. A. Predicting scientific impact based on h-index. *Scientometrics*, v. 114, n. 3, p. 993-1010, 2018. Doi: <https://doi.org/10.1007/s11192-017-2618-1>.
- Bai, X.; Zhang, F.; Lee, I. Predicting the citations of scholarly paper. *Journal of Informetrics*, v. 13, n. 1, p. 407-418, 2019. Doi: <https://doi.org/10.1016/j.joi.2019.01.010>.
- Chicco, D.; Jurman, G. The advantages of the Matthews Correlation Coefficient (MCC) over f1 score and accuracy in binary classification evaluation. *BMC Genomics*, v. 21, n. 1, p. 6, 2020. Doi: <https://doi.org/10.1186/s12864-019-6413-7>.
- Frietsch, R.; Gruber, S.; Bornmann, L. The definition of highly cited researchers: the effect of different approaches on the empirical outcome. *Scientometrics*, 2025. Doi: <https://doi.org/10.1007/s11192-024-05158-1>.
- Hand, D. J.; Christen, P.; Kirielle, N. F\*: an interpretable transformation of the F-measure. *Machine Learning*, v. 110, n. 3, p. 451-456, 2021. Doi: <https://doi.org/10.1007/s10994-021-05964-1>.
- Hirako, J.; Sasano, R.; Takeda, K. Realistic citation count prediction task for newly published papers. In: Vlachos, A.; Augenstein, I. (ed.). Findings of the Association for Computational Linguistics: EACL 2023. *Dubrovnik, Croatia: Association for Computational Linguistics*, 2023. p. 1131-1141. Doi: <https://doi.org/10.18653/v1/2023.findings-eacl.84>.
- Hirsch, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, v. 102, n. 46, p. 16569-16572, 2005. Doi: <https://doi.org/10.1073/pnas.0507655102>.
- Koltun, V.; Hafner, D. The h-index is no longer an effective correlate of scientific reputation. *Plos One*, v. 16, n. 6, p. 1-16, 2021. Doi: <https://doi.org/10.1371/journal.pone.0253397>.

- Manolopoulos, Y. et al. *Predicting the dynamics of research impact*. Cham: Springer International Publishing, 2021.
- Marques-Cruz, M. et al. Ten year citation prediction model for systematic reviews using early years citation data. *Scientometrics*, v. 129, n. 8, p. 4847-4862, 2024. Doi: <https://doi.org/10.1007/s11192-024-05105-0>.
- Momeni, F.; Mayr, P.; Dietze, S. Investigating the contribution of author- and publication-specific features to scholars' h-index prediction. *EPJ Data Science*, v. 12, n. 1, 2023. Doi: <https://doi.org/10.1140/epjds/s13688-023-00421-6>.
- Sinatra, R. et al. Quantifying the evolution of individual scientific impact. *Science*, v. 354, n. 6312, p. aaf5239, 2016. Doi: <https://doi.org/10.1126/science.aaf5239>.
- Teplitskiy, M. et al. How status of research papers affects the way they are read and cited. *Research Policy*, v. 51, n. 4, p. 104484, 2022. Doi: <https://doi.org/10.1016/j.respol.2022.104484>.
- Thai-Nghe, N.; Gantner, Z.; Schmidt-Thieme, L. Cost-sensitive learning methods for imbalanced data. In: The 2010 International Joint Conference on Neural Networks (IJCNN). [S.l.: s.n.], 2010. p. 1-8.
- Xia, W.; Li, T.; Li, C. A review of scientific impact prediction: tasks, features and methods. *Scientometrics*, v. 128, n. 1, p. 543-585, 2023. Doi: <https://doi.org/10.1007/s11192-022-04547-8>.
- Zhao, Q.; Feng, X. Utilizing citation network structure to predict paper citation counts: A Deep learning approach. *Journal of Informetrics*, v. 16, n. 1, p. 101235. 2022. Doi: <https://doi.org/10.1016/j.joi.2021.101235>.

## Collaborators

Conceptualization: I. C. DANTAS, A. A. BATISTA-JÚNIOR, L. R. COUTINHO, and J. P. MENA-CHALCO. Data preparation: I. C. DANTAS and A. A. BATISTA-JÚNIOR. Investigation: I. C. DANTAS, A. A. BATISTA-JÚNIOR, L. R. COUTINHO, and J. P. MENA-CHALCO. Methodology: I. C. DANTAS, A. A. BATISTA-JÚNIOR, L. R. COUTINHO, and J. P. MENA-CHALCO. Software: I. C. DANTAS and A. A. BATISTA-JÚNIOR. Supervision: L. R. COUTINHO and J. P. MENA-CHALCO. Validation: I. C. DANTAS, A. A. BATISTA-JÚNIOR, L. R. COUTINHO, and J. P. MENA-CHALCO. Writing-original draft: I. C. DANTAS, A. A. BATISTA-JÚNIOR, L. R. COUTINHO, and J. P. MENA-CHALCO. Writing-revision & editing: I. C. DANTAS, A. A. BATISTA-JÚNIOR, L. R. COUTINHO, and J. P. MENA-CHALCO.