International Journal of Information Science and Management
Vol. 23, No. 4, 2025, 63-87
DOI: https://doi.org/10.22034/ijism.2025.2044122.1630 / DOR: https://dor.isc.ac/dor/20.1001.1.20088302.2025.23.4.4.4

Original Research

Predicting the Economic Impact of Scientific Publications in Biotechnology Using Machine Learning

Ghasem Azadi Ahmadabadi
Assistant Prof., Policy Evaluation and Monitoring of Science, Technology, and Innovation Department, National Research Institute for Science Policy (NRISP), Tehran, Iran.
Corresponding Author: azadi@nrisp.ac.ir / azadi_gh@yahoo.com
ORCID iD: https://orcid.org/0000-0002-3610-2573

Hassan Bashiri
Assistant Prof., Department of Computer Engineering, Hamedan University of Technology, Hamedan, Iran.
bashiri@hut.ac.ir
ORCID iD: https://orcid.org/0000-0001-5429-5375

Received: 25 October 2024
Reviewed: 09 November 2024
Accepted: 14 July 2025

Abstract

The economic impact of research papers reveals the diffusion of information and its applicability to other technical fields. This research aims to predict the number of academic paper citations in patents. Papers gathered as a dataset for the study are the outputs of Iran's biotechnology field, indexed in the Scopus database from 2003 to 2024. To conduct the research, 15 indicators have been extracted for these articles in five categories: Journal, Altmetrics, Impact, Open Access, and Collaboration. We performed data processing, exploratory data analysis (EDA), machine learning modeling, and predictions using Python and libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn. The findings indicated that strong positive correlations are observed between the "Cite Score" and "SJR" indices, reflecting their related nature in evaluating journal impact. The "impact" category shows the strongest positive correlation with "patent information." The "journal" and "Altmetrics" categories show significant correlations, albeit to a lesser extent, indicating their complementary role in predicting economic impacts. Journal category indices, including SNIP, CiteScore, CiteScore percentile, SJR, and SJR percentile, exhibit a range of correlations with Patent citations. Altmetrics indices show a positive correlation with patent citations, which means that articles with higher visibility and engagement have a more significant impact on the patent literature. The results suggest that while machine learning is a powerful tool for predicting economic impact, further model refinement, feature selection, and more advanced techniques are necessary to achieve more accurate predictions. Considering the large gap between scientific papers and applied research in Iran's biotechnology field, is essential for managers and policymakers to identify and remove obstacles to the commercialization of scientific advancements.

Keywords: Scholarly Article Impact, Predicting Research Impact, Machine Learning, Economic Research Impact.

Introduction

Identifying relevant research and scholarly impacts is essential for the academic research communities and other stakeholders, such as governmental agencies and technological companies. In the research ecosystem where researchers publish their work results as research articles, citations are considered one of the general indicators of the articles' quality, relevance, and importance (Alohali, Fayed, Mesallam, Abdelsamad, Almuhawas & Hagr, 2022). Several features of research impact are difficult to measure. Impact can seldom be attributed to a single research study, but instead typically arises from a body of knowledge that encompasses multiple areas from various research studies (Morris, Wooding, & Grant, 2011). The domain of research impacts has predominantly been influenced by practical research policy demands, often lacking thorough and intentional consideration regarding the understanding and evaluation of research benefits to society (Williams & Grant, 2018). Research evaluation and scientific impacts, while often multidimensional, are frequently based on citations received by academic outputs (Radicchi, Weissman & Bollen, 2017). However, the scientometrics questions arise: which indicator (or mix) best predicts future impacts, and ii) How much is its predictive power? The answers to these inquiries vary depending on the time elapsed between the date of distribution and the estimation of gathered citations (Abramo, 2018). Researchers have employed various methods to assess a paper's scientific impact, but peer review and citation analysis are among the most significant. Over the past few decades, evaluating publications based on the number of citations they receive has become the gold standard for scientific evaluation (Akella, Alhoori, Kondamudi, Freeman, & Zhou, 2021).

Evaluating and predicting research impacts have attracted serious attention in scientific and academic fields in the past decades. The changes happen from one aspect to multiple aspects, from unstructured measures to structured measures (Figure 1).

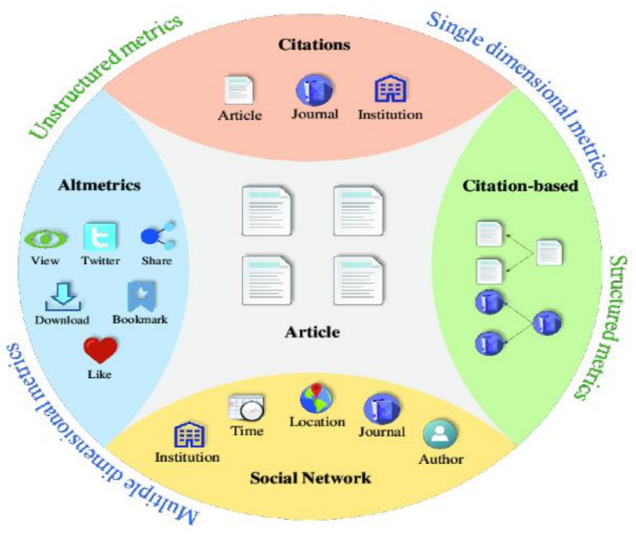

Figure 1: Methods and approach of evaluating and predicting paper impacts (Bai et al., 2017)

Based on Bai et al. (2017), citations are widely recognized as a key metric for assessing the impact of academic articles. However, this approach is limited to a singular dimension. In contrast, altmetrics offer a broader evaluation by incorporating data on downloads, views, shares, and citations, thereby providing a multidimensional assessment of article impact. The introduction of PageRank represents a significant advancement in the field of impact research, offering a systematic approach to quantifying article influence. Additionally, to ensure an objective evaluation of article impact and to accurately forecast its future significance, the application of machine learning and data mining techniques is essential. These methods facilitate the analysis of critical features within scholarly networks and enhance the efficacy of algorithms. Impact prediction is carried out using two main strategies. The first is a spiritual successor to Price's work, in which citation counts are intuitively statistically approximated based on an attachment model that prioritizes network development and actual research experience (Wang, Song & Barabási, 2013; Shen, Wang, Song & Barabási, 2014; Chen & Zhang, 2015). This approach leverages the inherent structure of scientific networks and the dynamics of citation behavior, offering insights rooted in traditional statistical modeling and domain expertise.

In contrast, the second strategy employs machine learning methods, such as deep feature engineering followed by supervised learning with regression models, to predict impact (Yan, Huang, Tang, Zhang & Li, 2012; Acuna, Allesina & Kording, 2012; Nezhadbiglari, Gonçalves & Almeida, 2016; Weihs & Etzioni, 2017; El Mohadab, Bouikhalene & Safi, 2019). This data-driven approach harnesses the power of algorithms to identify patterns and relationships within large datasets, providing a more automated and scalable solution for impact prediction. Together, these two strategies represent complementary perspectives: the first emphasizes interpretability and theoretical grounding, while the second prioritizes computational efficiency and adaptability to complex, high-dimensional data. Both approaches contribute valuable insights to the field of impact prediction, catering to different needs and contexts in research evaluation.

Predicting scientific article citations is crucial in guiding funding allocation, recruitment, and reward decisions. The models use multiple prediction features based on author, journal, and citations (Bai, Zhang & Lee, 2019). Funding agencies and researchers with limited time and resources are increasingly looking for models and measures to quantify the potential impact of collaborations or proposals (Yu, Yu, Li & Wang, 2014). In response to this growing demand, platforms like Elsevier have developed advanced tools and metrics to provide a more comprehensive evaluation of research impact. Elsevier, a world-leading supplier of scientific, technical, and medical information, reported the dispatch of unused measurements to help gauge the economic impact in SciVal. The new measures are aligned with SciVal’s mission to provide a wide range of measures to analyze and evaluate different dimensions of research. The release includes citations of patent articles as well as the incorporation of publication views data sourced from Scopus (Virtusnational, 2016).

“Economic Impact” encompasses the following metrics:

- Citing-Patents Counts: The number of patents citing scholarly outputs
- Patent-Cited Scholarly Outputs: The number of scholarly outputs that have been cited in patents
- Patent-Citations Counts: The number of patent citations received by the selected entity
- Patent-Citations per Scholarly Outputs: Average patent-citations per 1,000 scholarly outputs.

For evaluation research on economic impact, patent citations contain significant information. With thorough analysis, they can occasionally uncover hidden insights within the information flow between countries, universities, laboratories, and companies. Patent citation reveals the dissemination of information and its relevance across various technical domains, leading to the emergence of new technologies (Sharma & Tripathi, 2017). Patent citations provide perceptions of innovation pathways, the potential economic and technological outcomes, and the impact of research programs (Elsevier, 2024). Patent documents provide details regarding inventors, owners, the nations in which the inventions are protected, the technical domain of the invention, and citations of both academic publications and patent literature (Clarke, 2018). The primary relevant characteristics of patent citations are presented below:

- Patent data is accessible to the public and serves as a resource for statistical analysis aimed at assessing innovation, technological trends, and research and development activities, among other applications.
- Patent citations to academic publications signify a relationship between research and industry, highlighting original research as a fundamental contributor to innovation.
- Patents frequently cite research articles and other patents (Elsevier, 2019).

Thus, from a scientometric perspective, the economic impact of scientific research can be defined as the extent to which these researches are cited in patents or receives citations from patents. In fact, this citation exchange demonstrates how closely the research is tied to or influenced by technology, economics, and products. Therefore, it is essential to understand the extent to which published articles have had an economic impact and which indicators and components have been most significant in this regard. In our study, we tried to examine this issue and, as an example, on articles in the field of biotechnology that had economic impacts (Citing-Patents Counts, Patent-Cited Scholarly Outputs, Patent-Citations Counts, Patent-Citations per Scholarly Outputs).

Research Questions

The question remains concerning the potential to identify and forecast the effects of research by analyzing a combination of article attributes. Machine learning can be helpful because it has two main advantages. First, machine learning techniques enable the simultaneous analysis of multiple features, which can lead to significant discoveries. Second, machine learning methods allow modeling a target variable, such as the impact of an article, based on various characteristics, to make predictions about its scientific impact.

The purpose of this study is to utilize machine learning to examine the characteristics of Iranian scientific publications in biotechnology that were published between 2003 and 2024 and indexed in Scopus. In this work, we try to identify the combination of variables and indicators, including Journal (SNIP, Cite Score, Cite Score percentile, SJR, SJR percentile), Altmetrics (Field-Weighted View Impact, Views), Impact (Citations, Field-Weighted Citation Impact, Field-Citation Average, Outputs in Top Citation Percentiles, per percentile, Field-Weighted Outputs in Top Citation Percentiles, per percentile, Topic Prominence Percentile), Open Access, and Collaboration by different machine learning algorithms for assessment that best predicts citations in patents.

Literature Review

Finding and applying more efficient methods and models for discovering and predicting research impact, as well as scientific implications, will significantly increase the rate of uptake of scientific ideas. Azadi Ahmadabadi (2025) examines which machine algorithms can predict the scientific, social, and economic impact of scientific outputs. Machine learning algorithms such as multiple linear regression, nearest neighbor, decision trees, random forests, and gradient boosting were also used and evaluated as predictive models. Based on the obtained results, multivariate linear regression with a higher accuracy score and a lower standard deviation score could better predict the scientific, technological, and social impact of Iran's scientific outputs in biotechnology.

In their study, Gao, Liu, Pan & Wang (2024) focused on predicting the number of citations in the statistics field. They collected 55,024 academic articles published in 43 statistical journals between 2001 and 2018. Then they constructed multilayer networks from various aspects, including journal networks, author citation networks, co-citation networks, co-authorship networks, and keyword co-occurrence networks. Additionally, they identified 77 factors to predict the number of citations, comprising 22 traditional and 55 network-related factors. To consider the issues of under-inflated and over-dispersed citation numbers, a neural network model was designed to achieve high prediction accuracy. In addition, they adopted a leave-one-feature-out approach to examine the importance of these factors. The suggested neural network model achieved an average absolute error of 7.35, which is superior to other machine learning models. According to the researchers, this study provides a convenient guide to predict the citation numbers and can be extended to different research areas.

Zhang and Wu (2024) propose a pioneering method for predicting citation impact that employs multiple models designed for different research areas and integrates primary citation data. Their approach includes sample-based learning techniques to categorize articles into various research domains and other prediction models trained on the initial citation counts for articles in each domain. Their experimental findings confirm that the proposed prediction method, by using specific multi-domain models and primary citations, outperforms the four advanced basic methods in most cases and significantly increases the accuracy of citation impact prediction for diverse sets of academic articles.

Talaat and Gammel (2023) use two methods to investigate the effect of the number on the citation counts in research papers: (1) Pearson's correlation coefficient and (2) multiple regression. Empirical evidence suggests that co-authored papers achieve greater visibility and impact. They use multiple linear regression to predict the number of citations based on the number of authors, number of citations, location category, and year. The paramount originality of this study is the introduction of a practical predictor module that uses a probabilistic neural network to predict the citation numbers of the most effective items.

Alohali et al. (2022) suggested a machine learning model to predict the citation numbers of research outputs using different algorithms such as augmented decision trees, decision forests, and neural networks. Their results show that using neural networks and article abstracts provides the least root mean square error compared to using other algorithms, such as augmented decision trees or decision forests. They developed this algorithm using Microsoft Azure machine learning tools and an application using code-free programming technology that presents the dataset and predicts article citations using various algorithms.

Akella et al. (2021) utilized altmetrics to predict both short-term and long-term citations that a scientific output is likely to receive. They build different classification and regression models and evaluate their performances, identifying the most effective neural networks and ensemble models for these tasks. They also found that Mendelian readership was the most significant element for predicting primary citations, followed by other factors such as reader academic status (eg, student, postdoctoral fellow, professor), Twitter followers, online post length, number of authors, and Number mentioned on Twitter, Wikipedia, and in different countries.

Article impact was assessed by van der Zwaard, de Leeuw, Meerhoff, Bodine, and Knobbe (2020) for 4,531 publications with three impact measures: Altmetric attention scores, downloads, and citations. They built random forest regression models to predict paper impacts and identified papers with the highest impact (top 25% and top 10% for each impact metric), which were evaluated against a simple baseline method. Random forest models performed better than baseline models when predicting the influence of unseen articles. Also, random forest models predicted 25% and above 10% of influential articles with high accuracy. In addition, the random forest models showed the essential features of the paper. They confirmed that research impact can be predicted and better insight can be gained using machine learning and a combination of paper features.

Abrishmi and Aliakbari (2019) propose a new method to predict the long-term citations of an article based on citation numbers in the first years after publication. For training a citation prediction model, they used artificial neural networks, as a potent machine learning tool with recent applications in many fields, including image and text processing. Experimental tests show that their suggested method outperforms state-of-the-art methods in forecasting accuracy in annual forecasting and total number of citations.

The hypothesis was examined by Abramo, D'Angelo, and Felici (2019) that accurate predictions of long-term citation counts can be achieved by integrating a journal's initial citations with the impact factor of the journal in which it is published. The experiment is on a set of 123,128 WoS papers published by Italian authors, and the Fade is done from linear regression models. The average prediction accuracy is suitable for citation time intervals of more than two years, but decreases for less-cited publications and varies among different fields of study. They found that the role of the impact factor in the composition becomes negligible after two years of publication.

Stegehuis, Litvak, and Waltman (2015) employed a multiple regression model to predict future citation counts for a substantial collection of physics publications. Their analysis showed that both predictors (i.e., impact factor and initial citations) help accurately predict long-term citation impacts. They also analytically studied the behavior of multiple regression coefficients for upper quantiles of the distribution of citations.

Stern (2014) confirmed that received citations after the initial two years of publication account for over fifty percent of the variation in cumulative citations over an extended timeframe. While journal impact factors enhance both the predicted and actual correlation of journal articles based solely on citation data from 2006, this influence diminishes significantly in subsequent years.

By analyzing total citation data on the careers of 150,000 scientists, Mazloumian (2012) examined the fundamental assumption that citation counts serve as dependable indicators of future success. The results show that 1) among all citation indicators, annual citations at the time of forecasting are the best predictors of future citations, 2) future citations of a scientist's published articles can be accurately predicted (r2~0: 80 for a forecast year, Pv0:001), but 3) future citations of future works are hard to predict.

Therefore, measuring the impact of a scientific publication is a crucial, thus controversial matter. Overall, the literature analysis suggests that researchers have considered various approaches and characteristics when attempting to predict the scientific impact of their research.

Materials And Methods

This study aims to predict the economic impact of academic papers that emerge as patent citations. The research dataset is the outputs of Iran's biotechnology field, which are indexed in the Scopus database from 2003 through 2024. The data extraction date was 21 January 2024. The overall scholarly outputs were 14002, of which 384 records were cited in patents. We know these as influential articles in economics. The Scival analytical database, first introduced in 2009, was used to extract relevant data. The Scival database enables advanced and in-depth citation analyses based on the Scopus database. Using Scival, it is possible to evaluate the research performance of a researcher, organization, country, region, journal, research group, etc.

The research workflow is presented in Figure 2. The first step in this research was collecting information about the articles. For this purpose, the articles cited in the patents were identified in the ScienceDirect database, and their fifteen indicators were identified. The details of the indicators are given in the rest of the article. The second step was extracting the indicators related to the identified articles. Finally, the information related to the number of 384 extracted articles along with their indexes was collected in an Excel file. Two types of processing were performed on the data. One is the visual description of the data using the Pandas library in Python, and the other is the prediction of the economic impact of an article based on the extracted data using machine learning.

Figure 2: The research workflow

The 15 indicators were selected based on their established relevance in prior research on academic impact and economic influence (Abramo, 2018). Journal-based indicators assess publication visibility and prestige, while Altmetrics capture public engagement, and Impact indicators measure direct scholarly influence. Several fifteen indicators have been extracted for these articles in five categories: Journal, Altmetrics, Impact, Open Access, and Collaboration.

These indicators, along with their class titles, are:

Journal

1. SNIP (publication year): This index is calculated by weighting the citations based on the number of citations in the relevant subject areas (the citation potential of the database in the appropriate field) to correct the difference between the fields in terms of citation behavior and also in the amount of coverage in the database.

2. Cite Score (publication year) is a quantitative index used to measure the citation impact of scientific journals in the Scopus database. CiteScore indicates the average number of citations to articles in a scientific journal over a specific period. This index shows the relative importance of a journal compared to journals in the same field.
3. CiteScore percentile (publication year)
4. SJR (publication year): or SCImago index evaluates the impact of scientific articles based on the number of references and the importance of references. This term stands for SCImago Journal Rank and is used to calculate the impact factor of a scientific article.
5. SJR percentile (publication year)

Altmetrics

6. Field-Weighted View Impact: the average number of views per work, the weighted citation index of views
7. Views: the number of views, the output of the institution or university with the highest percentage of views.

Impact

8. Citations: number of citations received
9. Field-Weighted Citation Impact: weighted citation impact index at the field level, higher than one, indicates a better citation performance of the articles of that university compared to the global average of similar articles, and an index lower than one indicates a low performance of the articles of that university compared to the global average of similar articles.
10. Field-Citation Average: This metric represents the mean number of received citations by a collection of articles within a specific research domain during a particular year. It is determined by taking the total citations accrued by all papers classified under that research field and publication year and dividing it by the total number of documents in that category.
11. Outputs in Top Citation Percentiles: Publications among the top ten percentiles that have received the most citations worldwide. Superiority criteria help evaluate contributions in relation to influential and highly valuable articles in similar fields or disciplines. It can also be used to distinguish between researchers whose performance appears to be identical by other criteria such as scientific output, citations per output, or collaboration.
12. Field-Weighted Outputs in Top Citation Percentiles.
13. Topic Prominence Percentile: The prominence of a topic is a combination of the three criteria below and shows its mobility.
    - The number of citations in year n to articles published in year n and n-1.
    - Number of Scopus hits in year n to articles published in year n and n-1.
    - The average score of references (CiteScore) for year n

Open Access

14. Open Access: indicates open access to the full text of the article, which does not matter.

Collaboration

15. Country/Region: indicates that the article in question was published in collaboration with other countries, where the title of the country is not essential. It should not be only Iran; other countries are important.

To get a proper understanding of the data, we show five examples of the values of each data field.

Journal

1. SNIP (0.21, 048, 0.71, 062, 0.88)
2. Cite Score (0.2, 1, 4.2, 1.7, 3.6)
3. Cite Score percentile (91, 78, 47, 63, 41)
4. SJR (0.132, 0.23, 0.547, 0.415, 0.517)
5. SJR percentile (86, 74, 46, 56, 50)

Altmetrics

6. Field-Weighted View Impact (0.24, 0.68, 2.21, 0.52, 1.47)
7. Views (42, 15, 48, 21, 55)

Impact

8. Citations (0, 0, 1, 1, 1)
9. Field-Weighted Citation Impact (0, 0, 0.13, 0, 0.08)
10. Field-Citation Average (8.69, 11.6, 7.94, 31.22, 12.46)
11. Outputs in Top Citation Percentiles, per percentile (79, 79, 60, 70, 68)
12. Field-weighted outputs in Top Citation Percentiles, per percentile (68, 78, 69, 75, 74)
13. Topic Prominence Percentile (96.148, 97.45, 75.122, 98.818, 88.106)

Open Access

14. Open Access (Gold, Hybrid gold, Bronze |Green)

Collaboration

15. Country/Region (Iran, Switzerland |Iran, Taiwan| Iran, Canada)

Results

After collecting data from the articles and calculating the values of the introduced indicators, we utilized Python and libraries such as Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn for data processing, machine learning modeling, and prediction. A key consideration is that not all data was available during feature extraction, resulting in some empty cells in the prepared Excel file. Our calculations indicate a data density of 94.54% for the extracted data matrix. To address missing data in machine learning operations, empty cells were typically filled using the forward filling (ffill) method. This approach was selected to preserve data continuity while avoiding the potential bias introduced by random filling methods.

Figure 3, a heatmap of the correlation matrix, provides an overview of the relationships between the introduced indicators and their impact on the economic influence of articles. The color gradient reflects the strength and direction of the correlations, ranging from cold shades (indicating low or negative correlations) to warm shades (indicating strong positive correlations). As shown in Figure 3, 'CiteScore' and 'SJR' demonstrate a strong positive correlation, reinforcing their importance in journal evaluation. The 'impact' category exhibits the highest predictive power for patent citations, while 'journal' and 'Altmetrics' provide complementary insights. These findings are crucial for identifying the most valuable metrics in assessing the economic impact of articles.

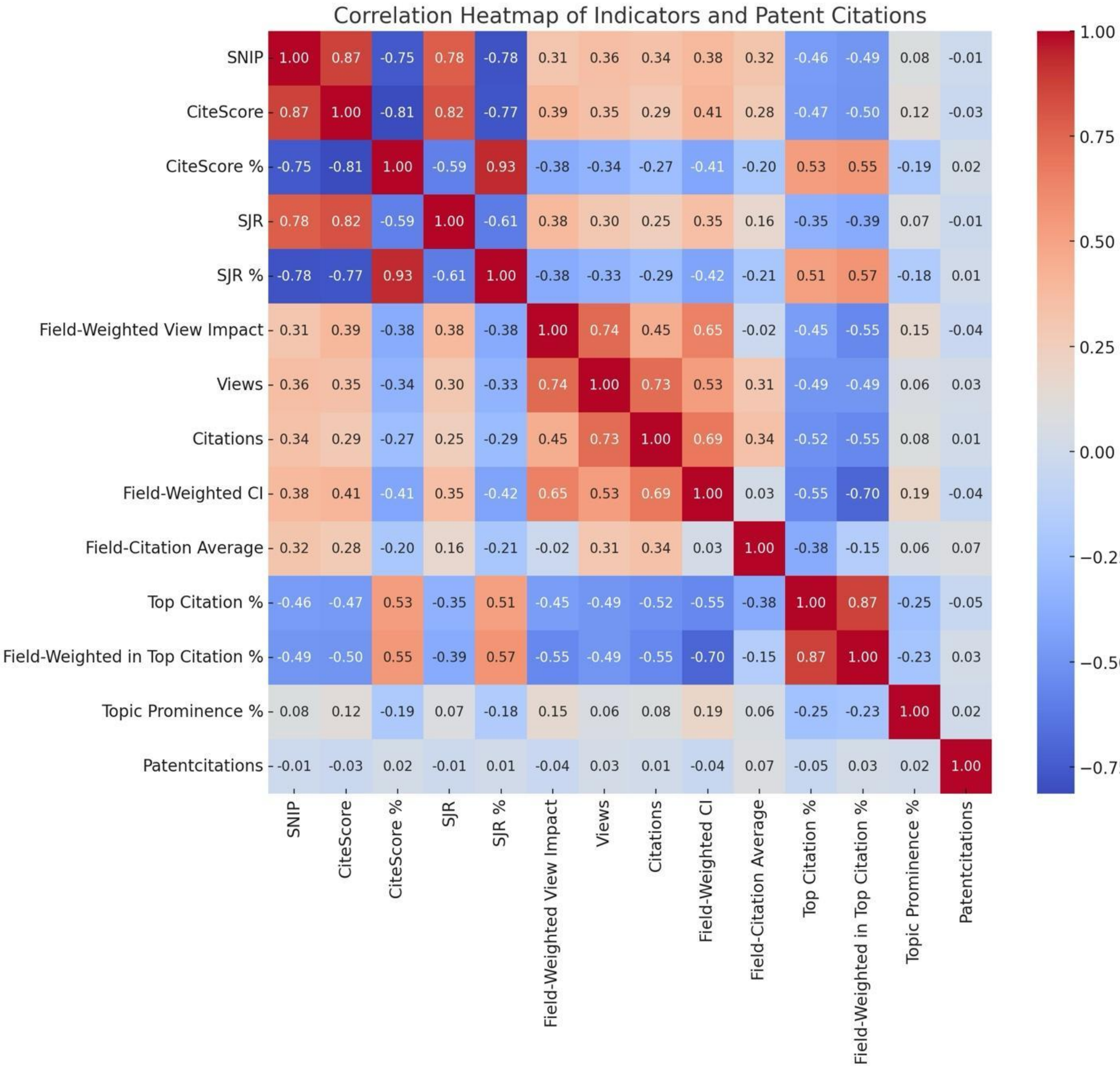

Figure 3: Correlation matrix heatmaps

To evaluate the importance of the fifteen indicators across five categories, we conducted an additional analysis, as shown in Figure 4. Notably, the last two indicators, related to the Open Access and Collaboration categories, were excluded from this analysis due to their non-numeric nature. Figure 4 illustrates the importance of each feature (indicator) in predicting patent citations. The features are ranked by importance, making it easy to identify which ones have the most significant impact on patent citations.

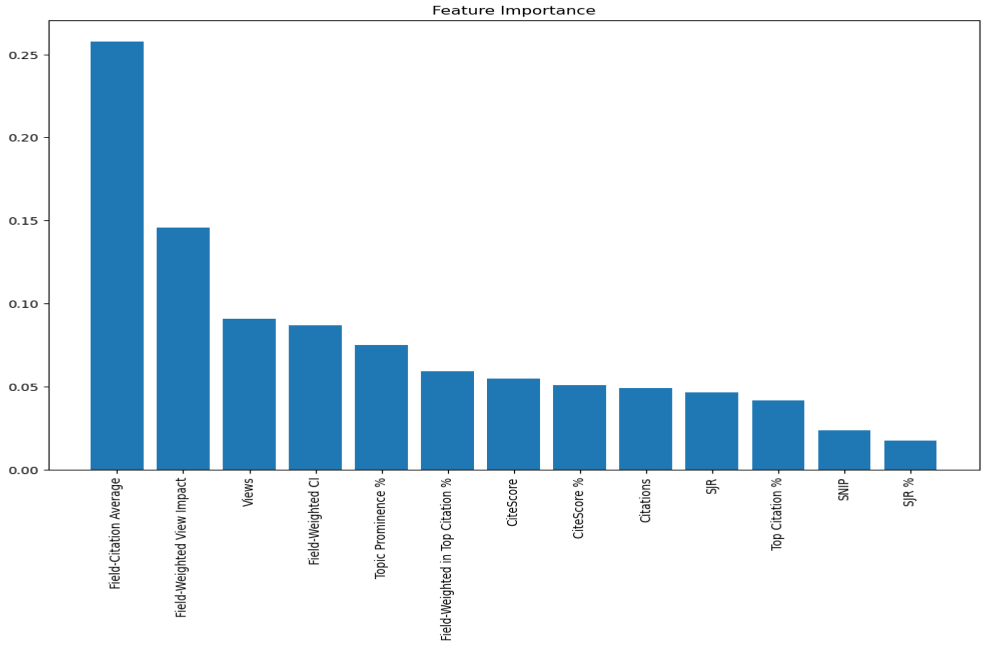

Figure 4: Importance of indicators from different categories

Another analysis focuses on the Mean Values of each category (Journal, Altmetrics, and Impact). This analysis allows us to calculate and compare the average impact of indicators across all articles. Figure 5 displays the Mean Values of thirteen indicators across three categories, while

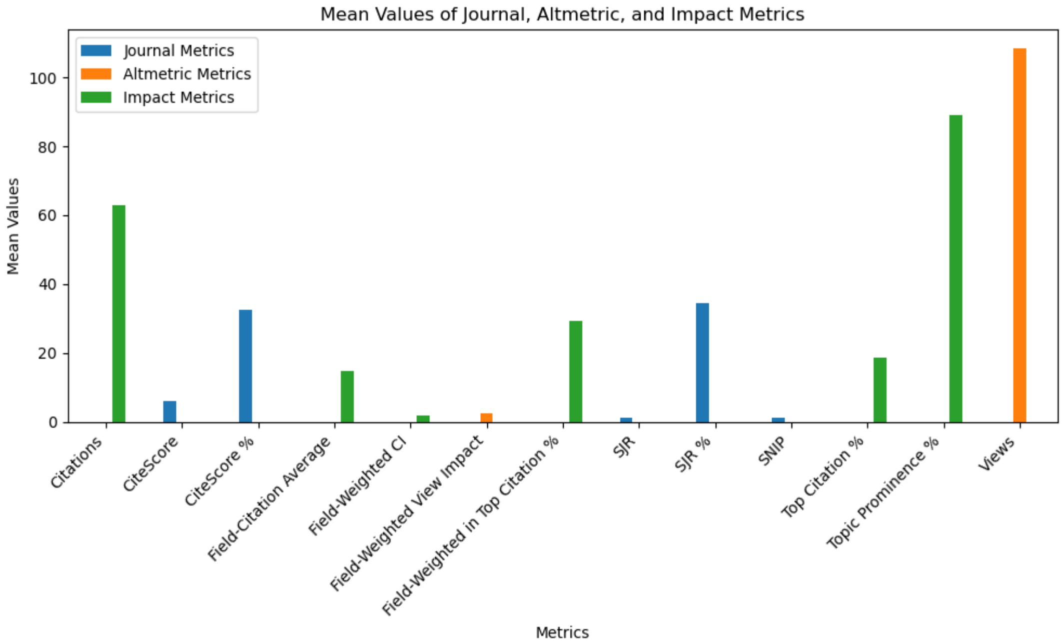

Figure 5: Mean Values for thirteen indicators belonging to three categories

Figure 6 presents the Mean Values of the categories themselves. By analyzing the mean values of indicators within each category, we can identify which categories tend to exhibit higher or lower average values. For instance, the Altmetrics category exhibits a higher mean value compared to the Journal and Impact categories, underscoring the significant role of

Altmetrics indicators in shaping the economic impact of articles.

The Mean Value for Journal Metrics reflects the overall performance of this category's indicators. Ultimately, this computed value represents the average ranking of articles within their respective journals, providing insight into their scientific credibility. In the Altmetrics category, indicators capture a paper's broader engagement beyond academia, including social media mentions, public discussions, and online interactions.

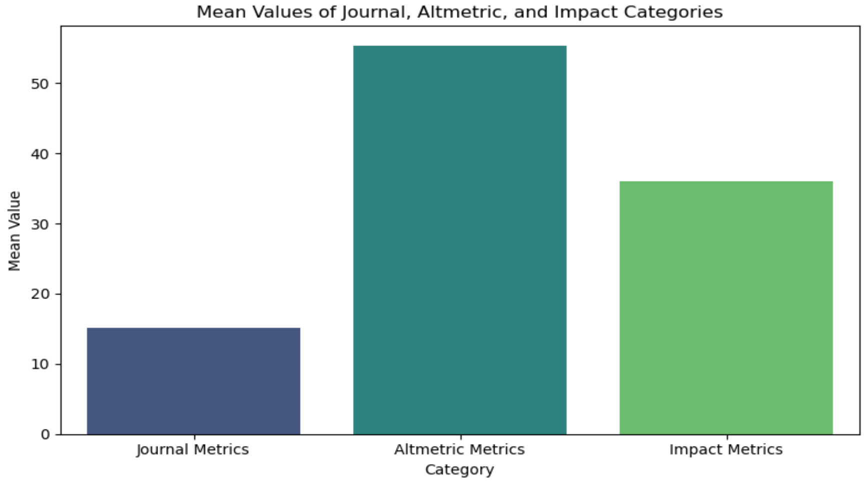

Figure 6: The mean values of the categories

Figure 7 presents a bar chart displaying the R^2 scores for three groups of indicators. The Open Access and Collaboration groups were excluded from this analysis due to their non-numeric values. Each bar indicates the extent to which the corresponding group of indicators explains variations in patent citations. The R^2 score measures the effectiveness of the independent variables (indicators) in predicting the dependent variable (patent citations). A higher R^2 score indicates greater influence of the corresponding group of indicators in predicting patent citations. The analysis reveals that Altmetrics and Impact indicators are the most critical factors influencing patent citations.

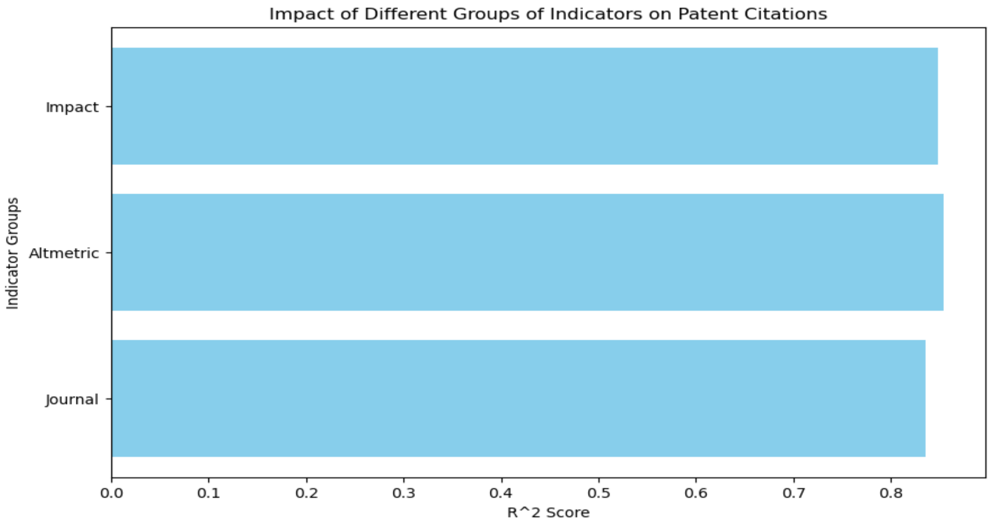

Figure 7: Impact of categories on patent citations

Figure 8 examines the influence of individual indicators, providing a more detailed perspective. Indicators with higher feature values are expected to have a stronger influence on driving patent citations.

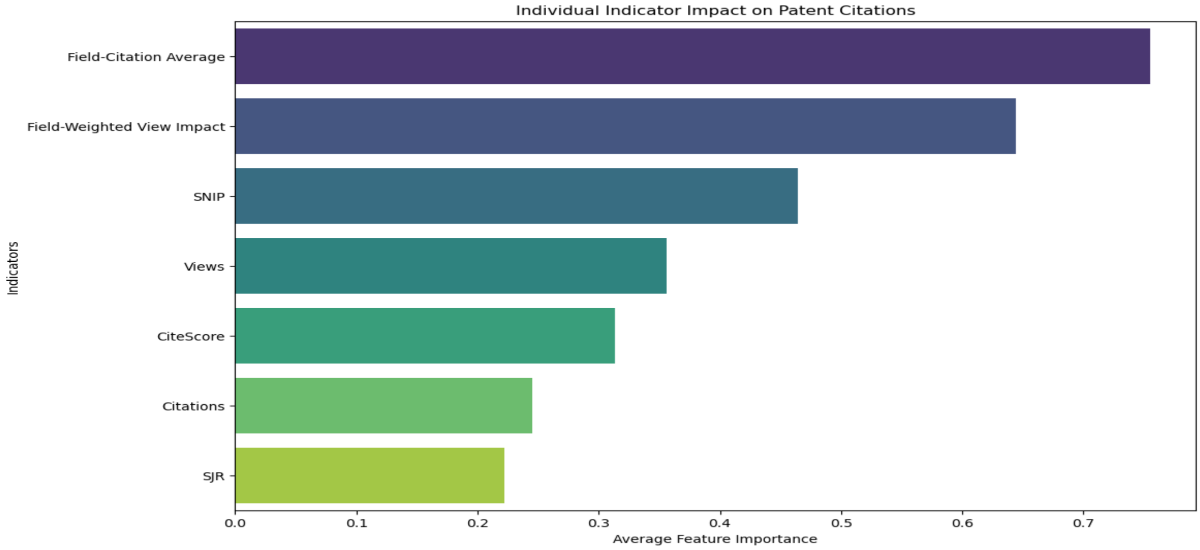

Figure 8: Indicators' impact on patent citations

In the machine learning section, the data set was divided into two parts: 80% for training and 20% for testing. For machine learning, we first applied a linear regression model to the data to predict the value of the dependent variable based on the independent variables. Here is the pseudo-code for our approach (Figure 9).

1. Load the dataset from an Excel file.
2. Fill missing values in the dataset with the mean of respective columns.
3. Separate the dataset into features (X) and the target variable (y).
4. Split the dataset into training and testing sets.
5. Train a Random Forest Regressor model on the training set.
6. Predict the target variable for the test set.
7. Calculate evaluation metrics (MSE, MAE, R-squared) for the model.
8. Print the evaluation metrics.
9. Define a function to predict patent citations for new input data.
10. Use the prediction function with example input values.
11. Plot the evaluation metrics using a bar plot.

Figure 9: Pseudocode of applying ML in the research

To predict the economic impact of research articles using machine learning, we followed the steps outlined in Figure 5. First, we preprocessed the data to handle missing values using appropriate methods. Next, we split the data into features—such as indices from the Journal, Altmetrics, and Impact categories—and target variables, which were patent citations. We trained the Random Forest Regressor model on this dataset to identify relationships between the features and the target variable. The model's performance was assessed using standard metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared (R^2).

These metrics provide insights into how effectively our model predicts the economic impact of new articles based on their indicators (Figure 10).

The Journal Metrics category, which includes indicators such as SNIP, CiteScore, CiteScore percentile, SJR, and SJR percentile, assesses the quality and impact of journals in which papers are published. These indicators reflect the journal's visibility and citation performance within academic communities, which in turn indirectly influence the economic impact of published articles.

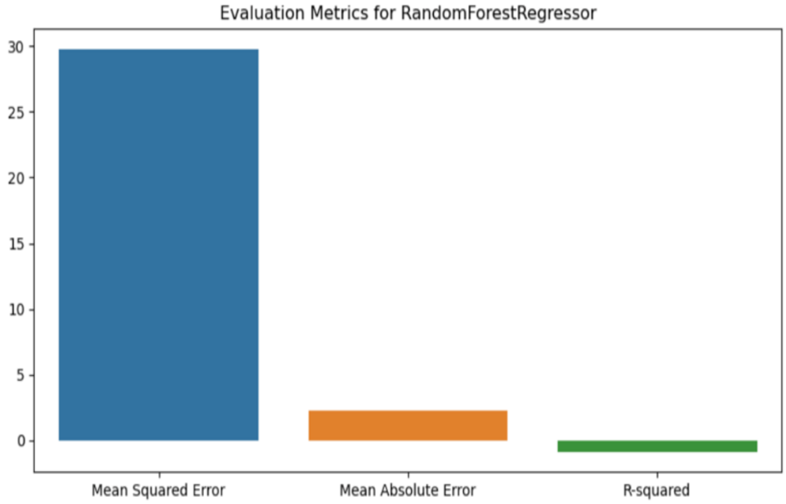

Figure 10: Evaluation metrics

Figure 11 presents a scatterplot illustrating the relationship between average journal metrics and patent citations. It is generally assumed that papers published in higher-impact journals are more likely to receive patent citations, suggesting a potential link between journal quality and the economic impact of research. However, the data presented in Figure 11 does not support this assumption. This can be attributed to the fact that highly cited journals often publish research with broader recognition. Nevertheless, other factors—such as the relevance of research to industry, its timeliness, and its potential for practical application in driving innovation—are equally critical in influencing patent citation counts.

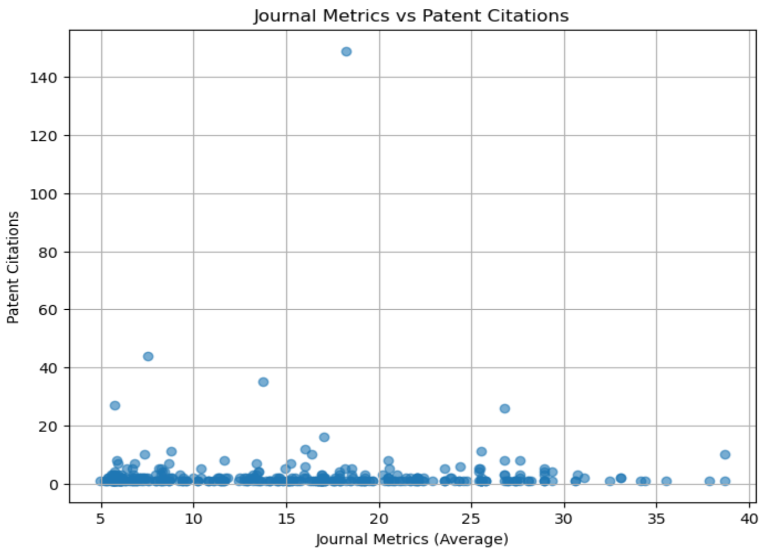

Figure 11: Relationship between Journal metrics and patent citations

Figures 12 and 13 also illustrate the relationship between Altmetrics, impact categories, and patent citations. Similar to the previous analysis, it is hypothesized that articles with higher Altmetrics scores or greater scientific impact are more frequently cited in patents. However, Figures 12 and 13 fail to demonstrate a clear linear relationship between these factors and patent citations. Consequently, other factors-such as timing, researcher-industry collaboration, research applicability, and project outcomes-play a significant role in determining the economic success of an article in terms of patent citations.

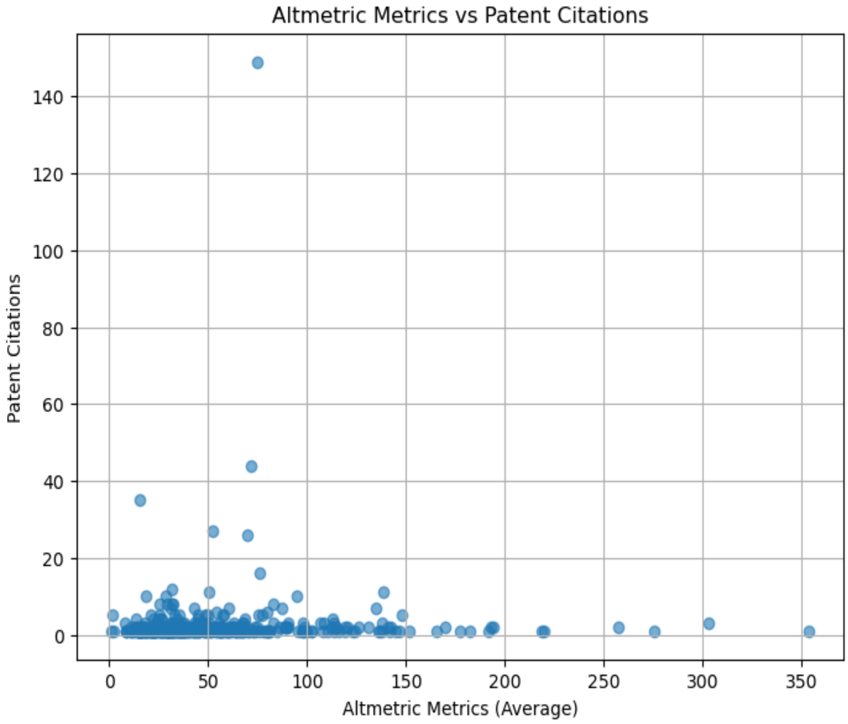

Figure 12: Relationship between altmetrics and patent citations

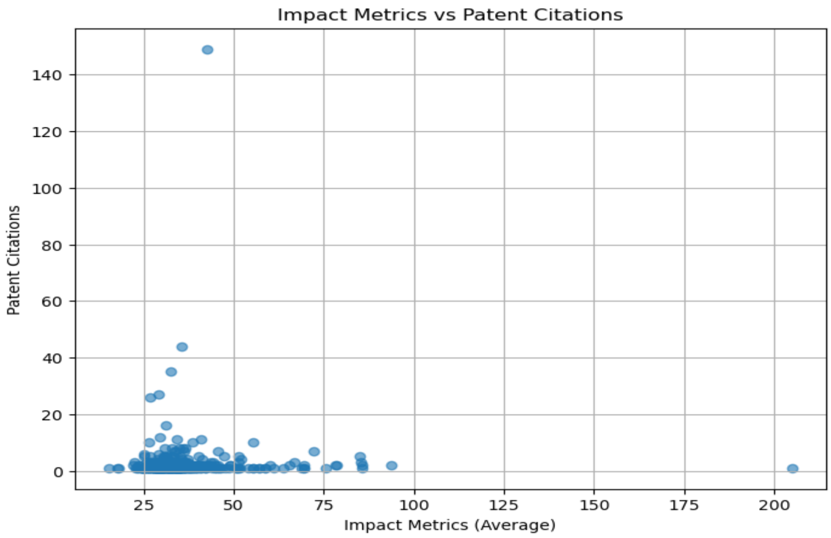

Figure 13: Relationship between impact metrics and patent citations

Discussion

Measuring the impact of research is a critical issue for universities and researchers across many nations. This study focuses on predicting the economic impact of academic papers within Iran's biotechnology field, specifically examining outputs with measurable economic contributions. To address this objective, we analyzed a comprehensive set of variables and indicators organized into three categories:

- Journal Metrics: Including SNIP, CiteScore, CiteScore Percentile, SJR, and SJR Percentile, which reflect the influence and prestige of journals. These metrics are widely used in academia to evaluate the quality and reach of scholarly publications, providing insights into how likely a paper is to be cited by other researchers.
- Altmetrics: Such as Field-Weighted View Impact and Views, capturing broader engagement with research outputs beyond traditional citations. Altmetrics offer a more dynamic perspective on research impact by tracking mentions on social media, policy documents, and news outlets, thereby reflecting societal relevance and public interest.
- Impact Metrics: Including Citations, Field-Weighted Citation Impact, Field-Citation Average, Outputs in Top Citation Percentiles (per percentile), Field-Weighted Outputs in Top Citation Percentiles (per percentile), and Topic Prominence Percentile, which assess the scholarly and societal impact of research. These indicators help identify high-performing papers that make significant contributions to their respective fields and have the potential to drive innovation and economic growth.

We used various machine learning algorithms to identify the most effective predictors of patent citations. Techniques such as decision trees, random forests, and neural networks were employed to uncover complex patterns and relationships within the dataset. The entire process—data processing, exploratory data analysis (EDA), machine learning modeling, and predictions—was conducted using Python and its robust libraries, including Pandas, NumPy, Matplotlib, Seaborn, and Scikit-Learn. These tools facilitated the efficient analysis and modeling of complex relationships within the dataset. Through the use of advanced computational methods, we aimed to provide actionable insights for policymakers, researchers, and funding agencies seeking to maximize the economic returns of scientific research.

The findings revealed strong positive correlations between the 'CiteScore' and 'SJR' indices, highlighting the strong relationship between these indices in assessing journal impact. These correlations suggest that CiteScore can serve as a reliable predictor of journal performance if accurate forecasting methods are developed. To achieve this, researchers must focus on creating robust yet straightforward models for predicting CiteScore, ensuring the feasibility and reliability of such predictions. A promising way to enhance forecasting accuracy is the integration of machine learning algorithms, capable of detecting non-linear relationships and complex patterns in citation data. For example, supervised learning models can be trained on historical citation trends to predict future CiteScore values more precisely. Furthermore, incorporating factors such as research community size, academic field, and publication type can further refine these predictions.

Further insights come from Stern (2014), who demonstrated that citations received after the first two years of publication can reliably predict long-term citation trends. This finding underscores the importance of monitoring citation trajectories over time, as early-stage metrics alone may not fully capture a paper's eventual influence. Similarly, Stegehuis et al. (2015) highlighted the critical role of early indicators, such as the journal impact factor and initial citation counts, in forecasting the sustained influence of scholarly works. Their analysis emphasizes that these early metrics are not only reflective of a paper's immediate reception but also predictive of its long-term impact.

It is worth noting that while journal-level metrics like CiteScore and SJR provide valuable insights, they should not overshadow article-level assessments. Individual papers may perform differently from the journals in which they are published, and focusing solely on journal metrics could lead to an incomplete understanding of research impact. Therefore, combining journal-level indicators with article-specific data, such as altmetrics or citation networks, can offer a more comprehensive evaluation of scholarly contributions.

Together, these findings underscore the importance of using early-stage data to predict the long-term impact of academic research. By developing robust predictive models and integrating multiple layers of data, researchers and policymakers can make better-informed decisions about funding allocations, strategic priorities, and the dissemination of high-impact research. Moreover, these efforts can contribute to a fairer and more transparent evaluation system, where the true value of scientific work is recognized beyond traditional metrics. Understanding future journal performance can significantly enhance decision-making processes. For academic institutions and nations, the quality of relevant journals serves as a key criterion for assessing their scientific contributions. This becomes especially crucial in highly competitive academic environments where institutions aim to strengthen their global standing and reputation. Publishing in high-quality, reputable journals not only demonstrates an institution's research excellence but also attracts funding, fosters collaborations, and draws top-tier faculty and students.

Moreover, authors and research teams often prioritize the quality of a journal when selecting the most suitable venue for their research. Key factors, such as the journal's impact factor, CiteScore, and reputation, significantly influence this decision. Publishing in high-impact journals enhances the visibility and credibility of research, amplifying its potential to affect both academic discourse and real-world applications. For early-career researchers, selecting the right journal is particularly crucial, as it can significantly impact their career trajectory. Publishing in reputable journals often serves as a key benchmark for evaluating a researcher's productivity and expertise, impacting tenure evaluations, grant opportunities, and professional acclaim. Additionally, anticipating future journal performance enables researchers to avoid submitting to journals that may decline in quality or relevance over time, ensuring their research reaches platforms with maximum impact and visibility.

At the national level, governments and funding agencies can utilize insights into journal performance to prioritize investments in high-potential research areas. By pinpointing journals with growing influence, policymakers can direct researchers toward fields and publications aligned with strategic goals, such as addressing societal challenges or fostering innovation in key industries. Ultimately, understanding future journal performance not only benefits individual researchers and institutions but also enhances the broader progress of science by optimizing the dissemination and impact of scholarly work. The 'impact' category exhibits the strongest positive correlation with 'patent information,' suggesting that citation-related indicators and field-weighted metrics are more predictive of a paper's economic impact. Key metrics, such as citation counts, Field-Weighted Citation Impact (FWCI), and top-percentile outputs, play a crucial role in evaluating a paper's contributions to both academic and practical domains. These metrics not only highlight the scholarly influence of research but also underscore its potential to drive innovation and generate economic value, especially when linked to patents or technological advancements.

Citations play several critical roles in scientific articles. They credit previous research, position the current study within the broader academic context, and provide evidence to substantiate claims. The number of citations a publication receives can indicate its utility, significance, or influence within the scientific community. However, citation counts alone do not fully capture a paper's impact. Other factors, including the quality of citing documents, the diversity of referencing fields, and altmetrics like social media mentions or policy influences, should be considered for a more comprehensive evaluation. According to Mazloumian (2012), annual citation counts at the time of prediction are the most reliable indicator of future citations. This finding highlights the significant predictive potential of early-stage citation data in forecasting long-term scholarly impact. For example, papers with a high initial citation count are more likely to maintain their relevance and influence over time. This insight is particularly valuable for researchers, funding agencies, and policymakers who aim to identify high-impact research early in its lifecycle.

Integrating traditional citation metrics with newer indicators, such as patent-related data or altmetrics, provides stakeholders with a more nuanced understanding of a paper's overall impact. This approach not only reflects the academic influence of research but also emphasizes its societal and economic contributions, aligning with the growing focus on measuring the broader impacts of science. Field-normalized citation impact reflects the average number of citations a publication receives, adjusted for factors such as field, year of publication, and document type (e.g., articles or reviews). This normalization ensures fair evaluation of research outputs by accounting for variations in citation practices across disciplines and over time. For instance, fields such as biotechnology typically exhibit higher citation rates compared to the humanities, underscoring the need for normalization to ensure equitable comparisons. Van der Zwaard et al. (2020) demonstrated that integrating article attributes with machine learning techniques can more effectively predict and understand research impact. By incorporating features such as author reputation, collaboration networks, and early-stage citation trends, predictive models can provide deeper insights into the long-term impact of scholarly works.

The 'journal' and 'Altmetrics' categories exhibit significant correlations, albeit to a lesser extent, highlighting their complementary roles in predicting economic impacts. While journal-based metrics, such as the impact factor, reflect the prestige and influence of the publishing platform, altmetrics capture broader societal engagement, including mentions on social media, in news outlets, or in policy documents. This dual perspective enables a more comprehensive assessment of a paper's reach and relevance, bridging the gap between academic and societal impact. Abramo et al. (2019) found that early-stage citation metrics, combined with the journal's impact factor, can predict citation counts effectively. Their findings revealed that the importance of the impact factor declines after two years, suggesting that early-stage indicators-such as initial citations-are more reliable predictors of long-term impact.

These insights underscore the value of adopting a multifaceted approach to research evaluation. By integrating traditional metrics such as journal impact factors with emerging indicators like altmetrics and machine learning-driven predictions, stakeholders can achieve a more comprehensive understanding of both academic and societal contributions. This integrated approach also aligns with the increasing focus on evaluating the broader impacts of research, including its economic, technological, and societal dimensions. Although traditional bibliometric indicators like citation counts and journal impact factors remain central to assessing research impact in biotechnology, altmetrics is emerging as a significant supplementary indicator of societal influence. Altmetrics encompasses a wide range of indicators, including social media mentions, policy citations, and news coverage, providing a broader perspective on an article's reach and impact. Integrating altmetrics with traditional metrics offers valuable insights into the most effective measures for assessing the economic impact of research outputs. By analyzing these indicators, researchers and institutions can gain deeper insights into the broader impact of their work, transcending traditional academic metrics. Academic stakeholders must acknowledge the growing significance of alternative metrics in evaluating the societal contributions of scholarly articles.

Indices such as SNIP, CiteScore, CiteScore Percentile, SJR, and SJR Percentile show varying degrees of correlation with patent citations. Notably, SNIP and CiteScore exhibit a moderate positive correlation, indicating that articles published in journals with higher citation impact are more likely to be cited in patents. This relationship highlights the critical role of journal quality in enhancing both the visibility and practical relevance of research outputs. Among these metrics, Source Normalized Impact per Paper (SNIP) stands out for its ability to account for variations in citation practices across scientific fields. SNIP assesses contextual citation impact by normalizing citations based on the total citation volume within a specific subject area. In fields with fewer citations, a single citation holds more weight, whereas in disciplines with higher citation rates, individual citations are proportionally less significant. This normalization renders SNIP especially valuable for interdisciplinary research, where citation patterns can vary significantly across contributing fields. As an inherently multidisciplinary field, biotechnology serves as a prime example of how SNIP facilitates fair and accurate comparisons of research impact. By accounting for differences in citation behavior across disciplines, SNIP enables more precise evaluations of citation impact, helping researchers and policymakers better assess the influence of scholarly work in diverse contexts, including those related to patents.

However, the SJR percentile exhibits a negative correlation, suggesting that higher journal rankings do not necessarily translate into increased patent citations. Altmetrics indices demonstrate a positive correlation with patent citations, indicating that articles with greater visibility and engagement are more likely to influence patent literature. In general, patent citations occur infrequently, with a notable bias toward older publications due to their extended time to gain visibility and recognition. This bias stems from older articles having had more time to gain visibility and be featured in various media, increasing their likelihood of being cited in patents. Akella et al. (2021) showed that altmetrics can effectively predict both short-term and long-term citation trends. Their study highlighted Mendeley readership as the most critical factor in forecasting future citations, underscoring its role as a key indicator of research impact. Building on these insights, Alchokr, Haider, Shakeel, Leich, Saake, and Krüger (2023) utilized a machine learning framework to confirm the effectiveness of altmetrics in predicting future citation counts. Collectively, these studies underscore the growing importance of altmetrics as a complementary tool for evaluating the impact of scholarly work, particularly in contexts such as patent citations, where traditional metrics may be insufficient.

Altmetrics track the visibility of scholarly articles by monitoring their presence across diverse platforms, including social networks, mainstream media, science blogs, policy documents, patents, Wikipedia entries, peer review platforms, F1000, syllabi, Stack Exchange sites, and YouTube. By consolidating data from these varied sources, Altmetrics provide an early signal of an article's potential impact, allowing researchers to assess its reach and influence before traditional citation metrics become available. Articles from prestigious institutions and top-tier journals often receive more attention, reflecting inherent biases in academic visibility. However, Altmetrics can help overcome these biases by highlighting engagement beyond traditional measures, such as social media traction or policy document references. For example, articles gaining traction on social media or being cited in policy documents may achieve broader societal impact, irrespective of their institutional origin.

By adopting diverse strategies, such as engaging with social media, collaborating with science communicators, or contributing to policy discussions, and evaluating their effectiveness through Altmetrics, researchers can devise innovative ways to enhance the visibility and prominence of their work. These efforts not only broaden the reach of scientific research but also ensure its relevance and applicability across diverse contexts. Impact indicators include citations, Field-Weighted Citation Impact (FWCI), field citation average, top citation percentiles, and field-weighted outputs in top citation percentiles. Among these, the number of citations and Field-Weighted Citation Impact exhibit the strongest positive correlation with patent citations. This highlights the critical role of an article's citation performance, particularly its prominence in highly cited percentiles, in determining its influence on patents. Field-Weighted Citation Impact is especially valuable as it normalizes citation counts based on field-specific citation practices, providing a more equitable measure of an article's impact. Similarly, an article's placement in top citation percentiles underscores its exceptional contribution to its field, making it more likely to attract attention from patent developers and industry practitioners. These findings emphasize the importance of leveraging citation-based metrics to identify high-impact research with potential applications in innovation and technology development.

Although the machine learning model provided insights, its performance was suboptimal (MSE = 29.78, R^2 = -0.87). The model's weaknesses can be attributed to dataset limitations and the need for feature engineering. Future studies should refine variable selection and explore advanced modeling techniques. A negative R-squared value indicates that our model performs worse than a simple mean prediction, suggesting that it cannot effectively capture underlying patterns in the data. The predicted economic impact for several sample inputs also highlights discrepancies, with one prediction (113.32) being significantly higher than the others, suggesting potentially remote limits on the model's ability to generalize. These results indicate that machine learning is a powerful tool for predicting economic impact. Also, van der Zwaard et al (2020) confirmed that research impact can be expected and better understood using an aggregate of article attributions and machine learning. Azadi Ahmadabadi (2025) concluded that multivariate linear regression could better predict the scientific, technological, and social impact of scientific outputs. In the field of machine learning, multivariate regression is frequently employed as a supervised algorithm. This powerful technique serves as an effective means of forecasting the behavior of dependent variables by analyzing several independent variables. Talaat and Gamel (2023) approved using multiple linear regression to predict the citation numbers.

Conclusion

Citation counts, which measure the frequency with which other academic works reference an article, are widely regarded as the most relevant metric for evaluating the impact of scholarly articles, authors, journals, institutions, and countries (Bu, Lu, Wu, Chen, & Huang, 2021). This prominence stems from the assumption that citations reflect the perceived value or relevance of a work within the scientific community. However, it is essential to recognize that citation counts are not a perfect measure of academic impact. They can be influenced by factors such as the size of the research field, publication practices, and even biases in educational networks. Therefore, while citation counts provide valuable insights, they should ideally be complemented by other metrics, such as altmetrics or qualitative assessments, to capture a more holistic view of research impact. Predicting the future citation impact of academic papers provides significant advantages for various stakeholders within the research ecosystem. For publishers, predictive models can serve as a valuable tool in the evaluation process, enabling them to prioritize high-impact articles for promotion or inclusion in prestigious journals. For authors, understanding the factors that contribute to higher citation rates can guide improvements in the quality and presentation of their work, thereby increasing its visibility and likelihood of being cited. Additionally, funding agencies and institutions can use these predictions to allocate resources more effectively, supporting projects with the most significant potential for long-term impact. However, it is crucial to approach these predictions with caution, as over-reliance on quantitative measures like citations may inadvertently incentivize practices such as "citation gaming" or favoring trendy topics over foundational research (Szomszor, Pendlebury & Adams, 2020).

In this study, we tested a machine learning model to predict the citations of research papers in patents. By focusing on patent citations, the study highlights the intersection between academic research and its practical applications in innovation and industry. This approach underscores the growing importance of measuring not only academic impact but also economic and societal contributions. Indicators were extracted for the analyzed articles in three categories: Journal (e.g., journal impact factor), Altmetrics (e.g., social media mentions, news coverage), and Impact (e.g., author reputation, collaboration networks). These categories reflect a multi-dimensional view of research impact, acknowledging that citations alone do not fully capture the value of scholarly work. These results suggest that while machine learning is a powerful tool for predicting economic impact, further model refinement, feature selection, and more advanced techniques are necessary to achieve more accurate predictions. It is also worth noting that machine learning models, despite their predictive power, often operate as "black boxes," making it challenging to interpret how specific features influence outcomes. In scientometric research, where transparency and explainability are crucial, striking a balance between predictive accuracy and interpretability is essential. Collaborations between data scientists and domain experts can help bridge this gap, ensuring that the models not only perform well but also provide meaningful insights into the mechanisms driving research impact. The integration of machine learning into scientometrics represents a significant step forward in our ability to quantify and predict research impact. However, it is essential to remain mindful of the ethical and practical implications of relying heavily on algorithmic predictions. For example, overemphasis on citation-based metrics may disadvantage certain types of research, such as theoretical or interdisciplinary work, that may have profound but delayed impacts. As the field evolves, it will be critical to develop frameworks that combine quantitative predictions with qualitative evaluations, fostering a more equitable and comprehensive approach to assessing the value of scientific contributions.

One of the main limitations of our study is the small size of the dataset, which only includes 382 articles. This limited sample size can hinder the model's ability to generalize and accurately capture the underlying relationships between the characteristics and the target variable. With a small dataset, the risk of overfitting increases, meaning the model may perform exceptionally well on the training data but poorly on unseen data. Overfitting occurs when the model learns noise or specific details of the training set rather than generalizable patterns, which is particularly problematic in scientometric studies where citation behaviors can vary significantly across disciplines and contexts. To mitigate this issue, techniques such as cross-validation, regularization, or ensemble methods could be employed in future iterations of the model.

To improve the strength and accuracy of our predictions, we will seek to expand the dataset in future research. A larger and more diverse dataset would allow the model to learn from a broader range of citation patterns, enhancing its robustness and applicability across different fields. In addition, we will explore other, more complex machine learning algorithms, such as gradient boosting machines or neural networks, which are better suited for handling non-linear relationships and high-dimensional data. These advanced techniques could uncover hidden patterns and interactions that simpler models might miss. Furthermore, refining the feature selection process—by identifying the most relevant predictors and eliminating redundant variables—can reduce overfitting and improve interpretability. For instance, incorporating domain-specific knowledge to select features that align with the dynamics of citation behavior in patents could yield more accurate and actionable insights.

Based on the findings of this study, we have three categories of suggestions for researchers in Scientometrics, biotechnology scientists, and science policymakers.

We recommend that other researchers in Scientometrics conduct a separate study using data from different fields and compare their results with those of this research. We suggest that researchers use other machine learning algorithms for predicting the impact of patent citations of scientific papers. Suggestions for improving the Model:

1. Increasing the Volume of Data: To enhance the model's performance, researchers suggest increasing the volume of data in future studies. This can help the model identify deeper and more accurate patterns.

2. Using More Advanced Algorithms: The text recommends utilizing more complex machine learning algorithms, such as Gradient Boosting Machines or Neural Networks. These algorithms are capable of better modeling non-linear and complex relationships between features and the target variable.

3. Feature Selection: Optimizing the feature set can also help reduce the effect of Overfitting and improve prediction accuracy.

It is recommended that researchers focus on technological fields for their scientific work.

Given the positive correlation between online visit counts and indicators of scientific quality and effectiveness, researchers are encouraged to participate in scientific social networks actively. Sharing their research findings through these platforms can significantly enhance the visibility and impact of their work.

For policymakers in science and technology, fostering scientific development in Iran requires addressing both quantitative and qualitative growth in scientific publications. However, beyond sheer numbers, the country's research system must prioritize generating tangible economic value, such as through innovations that lead to patents or commercial applications. One notable challenge is the limited number of patent-related publications, which highlights the gap between theoretical research and practical outcomes. This gap is particularly evident in the biotechnology sector, where the translation of scientific findings into marketable innovations remains limited. To address this, managers and policymakers must identify and remove obstacles to the commercialization of scientific achievements. Potential strategies include providing incentives for applied research, strengthening collaborations between academia and industry, and establishing clear pathways for translating discoveries into economic value. Bridging this gap is crucial for advancing both scientific and economic development in Iran.

Acknowledgments

This article is based on a segment of the approved research project titled Evaluating the Impacts of Scientific Outputs: A Case Study of Iran's Biotechnology Field, conducted at the National Research Institute for Science Policy.

References

Abramo, G. (2018). Revisiting the scientometric conceptualization of impact and its measurement. *Journal of Informetrics,* 12(3), 590-597. https://doi.org/10.1016/j.joi.2018.05.001

Abramo, G., D'Angelo, C.A. & Felici, G. (2019). Predicting long-term publication impact through a combination of early citations and journal impact factor. *Journal of Informetrics*, 13(1), 32-49. https://doi.org/10.1016/j.joi.2018.11.003

Abrishami, A. & Aliakbary, S. (2019). Predicting citation counts based on deep neural network learning techniques. *Journal of Informetrics*, 13(2), 485-499. https://doi.org/10.1016/j.joi.2019.02.011

Acuna, D. E., Allesina, S. & Kording, K. P. (2012). Predicting scientific success. *Nature*, 489(7415), 201-202. https://doi.org/10.1038/489201a

Akella, A. P., Alhoori, H., Kondamudi, P. R., Freeman, C. & Zhou, H. (2021). Early indicators of scientific impact: Predicting citations with Altmetrics. *Journal of Informetrics*, 15(2), 101128. https://doi.org/10.1016/j.joi.2020.101128

Alchokr, R., Haider, R., Shakeel, Y., Leich, T., Saake, G. & Krüger, J. (2023). Forecasting Publication's Success Using Machine Learning. In *International Workshop on Bibliometric-Enhanced Information (BIR)* (pp. 1-13). CEUR-WS. org. Retrieved from https://jacobkrueger.github.io/assets/papers/Alchokr2023ForcastingSuccess.pdf

Alohali, Y. A., Fayed, M. S., Mesallam, T., Abdelsamad, Y., Almuhawas, F. & Hagr, A. (2022). A machine learning model to predict citation counts of scientific papers in the otology field. *BioMed Research International*, 2022(1), 2239152. https://doi.org/10.1155/2022/2239152

Azadi Ahmadabadi, G. (2025). Predicting scientific research impacts in biotechnology by machine learning algorithms. *Scientometrics Research Journal*, 11(1) 1-24. https://doi.org/10.22070/rsci.2024.18868.1719 [in Persian]

Bai, X., Liu, H., Zhang, F., Ning, Z., Kong, X., Lee, I., & Xia, F. (2017). An overview on evaluating and predicting scholarly article impact. *Information*, 8(3), 73. https://doi.org/10.48550/arXiv.2008.03867

Bai, X., Zhang, F. & Lee, I. (2019). Predicting the citations of a scholarly paper. *Journal of Informetrics*, 13(1), 407-418. https://doi.org/10.1016/j.joi.2019.01.010

Bu, Y., Lu, W., Wu, Y., Chen, H. & Huang, Y. (2021). How wide is the citation impact of scientific publications? A cross-disciplinary and large-scale analysis. *Information Processing & Management*, 58(1), 102429. https://doi.org/10.1016/j.ipm.2020.102429

Chen, J., & Zhang, C. (2015, July). Predicting citation counts of papers. In *2015, IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI CC)* (pp. 434-440). IEEE. Beijing, China. https://doi.org/10.1109/ICCI-CC.2015.7259421

Clarke, N. S. (2018). The basics of patent searching. *World Patent Information*, 54, S4-S10. https://doi.org/10.1016/j.wpi.2017.02.006

El Mohadab, M., Bouikhalene, B. & Safi, S. (2019). Predicting rank for scientific research papers using supervised learning. *Applied Computing and Informatics*, 15(2), 182-190. https://doi.org/10.1016/j.aci.2018.02.002

Elsevier (2019). *Research metrics guidebook*. Retrieved from https://elsevier.widen.net/s/chpzk57rqk/acad_rl_elsevierresearchmetricsbook_web

Elsevier (2024). *SciVal impact*. Retrieved from https://www.elsevier.com/products/scival/impact

Gao, T., Liu, J., Pan, R. & Wang, H. (2024). Citation counts prediction of statistical publications based on multi-layer academic networks via a neural network model. *Expert Systems with Applications*, 238, 121634. https://doi.org/10.1016/j.eswa.2023.121634

Mazloumian, A. (2012). Predicting scholars' scientific impact. PLoS ONE, 7(11), e49246. https://doi.org/10.1371/journal.pone.0049246

Morris, Z., Wooding, S. & Grant, J. (2011). The answer is 17 years. What is the question: Understanding time lags in translational research? *Journal of the Royal Society of Medicine*, 104(12), 510–520. https://doi.org/10.1258/jrsm.2011.110180

Nezhadbiglari, M., Gonçalves, M. A. & Almeida, J. M. (2016, June). Early prediction of scholar popularity. In *Proceedings of the 16th ACM/IEEE-CS on Joint Conference on Digital Libraries* (pp. 181-190). https://doi.org/10.1145/2910896.2910905

Radicchi, F., Weissman, A. & Bollen, J. (2017). Quantifying the perceived impact of scientific publications. *Journal of Informetrics*, 11(3), 704-712. https://doi.org/10.1016/J.JOI.2017.05.010

Sharma, P. & Tripathi, R. C. (2017). Patent citation: A technique for measuring the knowledge flow of information and innovation. *World Patent Information*, 51, 31-42. https://doi.org/10.1016/j.wpi.2017.11.002

Shen, H., Wang, D., Song, C. & Barabási, A. L. (2014, June). Modeling and predicting popularity dynamics via reinforced Poisson processes. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 28, No. 1). https://doi.org/10.1609/aaai.v28i1.8739

Stegehuis, C., Litvak, N. & Waltman, L. (2015). Predicting the long-term citation impact of recent publications. *Journal of Informetrics*, 9(3), 642-657. http://dx.doi.org/10.1016/j.joi.2015.06.005

Stern, D. I. (2014). High-ranked social science journal articles can be identified from early citation information. *PLoS One*, 9(11), e112520. http://dx.doi.org/10.1371/journal.pone.0112520

Szomszor, M., Pendlebury, D. A. & Adams, J. (2020). How much is too much? The difference between research influence and self-citation excess. *Scientometrics*, 123(2), 1119-1147. https://doi.org/10.1007/s11192-020-03417-5

Talaat, F. M., & Gamel, S. A. (2023). Predicting the impact of the number of authors on the number of citations of research publications based on neural networks. *Journal of Ambient Intelligence and Humanized Computing*, 14(7), 8499-8508. https://doi.org/10.1007/s12652-022-03882-1

van der Zwaard, S., de Leeuw, A. W., Meerhoff, L. R. A., Bodine, S. C. & Knobbe, A. (2020). Articles with impact: Insights into 10 years of research with machine learning. *Journal of Applied Physiology*, 129(4), 967-979. http://dx.doi.org/10.1152/japplphysiol.00489.2020

Virtusnational (2016). Elsevier announces the launch of new metrics on Scival to help institutions measure the economic impact of their research. Retrieved from https://www.prnewswire.com/news-releases/elsevier-announces-the-launch-of-new-metrics-on-scival-to-help-institutions-measure-the-economic-impact-of-their-research-570508761.html

Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science*, 342(6154), 127-132. https://doi.org/10.1126/science.1237825

Weihs, L. & Etzioni, O. (2017, June). Learning to predict citation-based impact measures. In *2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL)* (pp. 1-10). IEEE. https://doi.org/10.1109/JCDL.2017.7991559

Williams, K. & Grant, J. (2018). A comparative review of how the policy and procedures to assess research impact evolved in Australia and the UK. *Research Evaluation*, 27(2), 93–105. https://doi.org/10.1093/reseval/rvx042

Yan, R., Huang, C., Tang, J., Zhang, Y., & Li, X. (2012, June). To better stand on the shoulders of giants. In *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries* (pp. 51-60). https://doi.org/10.1145/2232817.2232831

Yu, T., Yu, G., Li, P. Y. & Wang, L. (2014). Citation Impact Prediction for Scientific Papers Using Stepwise Regression Analysis. *Scientometrics*, 101(2), 1233-1252. https://doi.org/10.1007/s11192-014-1279-6

Zhang, F. & Wu, S. (2024). Predicting citation impact of academic papers across research areas using multiple models and early citations. *Scientometrics*, 129(7), 4137-4166. https://doi.org/10.1007/s11192-024-05086-0