Citation Evidence Report

EB-1B Petition — Outstanding Professor or Researcher

8 CFR § 204.5(i)(3) · Authorship + Original Contributions

Philip S. Yu

Professor of Computer Science, University of Illinons at Chicago

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to the 8 CFR § 204.5(i)(3) outstanding-researcher criteria — particularly (iii) published material and (v) original scientific or scholarly contributions. It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

54	54	6	217
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

88.9% independent of 54 classified citing papers

Citation type	Count
Independent	48
Self-citation	0
Co-author	6
Same-institution	0

0 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher provided a foundational database-centric overview of data mining, establishing a seminal framework that has been widely adopted by the independent academic community.

CLAIM: The researcher's core contribution is the publication of a seminal overview of data mining from a database perspective, published in IEEE Transactions on Knowledge and Data Engineering in 2002. This work serves as the primary anchor for this line of research, standing alone without direct follow-up papers by the same author in the provided dataset.

ORIGINALITY: The title suggests the work addressed a critical need to contextualize data mining techniques within the established domain of database systems. By framing data mining through a database lens, the researcher likely provided a structured, systematic perspective that distinguished these emerging analytical methods from traditional database operations, offering a novel synthesis for the field at that time.

SIGNIFICANCE: The work has achieved substantial impact, evidenced by over 4,000 citations. Notably, citation analysis reveals that 100% of the classified citing papers originate from independent researchers, indicating that the contribution has been widely recognized and utilized by the broader scientific community rather than just the researcher's immediate circle. This high level of independent uptake underscores the work's status as a foundational reference in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 9

CORE PAPER

Data Mining: An Overview from a Database Perspective

2002 · IEEE Transactions on Knowledge and Data Engineering · 4,085 citations (GS)

Field-normalised: 2,621 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2002 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	A Tutorial on Environment-Aware Communications via Channel Knowledge Map for 6G (2025)	EURECOM, Southeast University, The Chinese University of Hong Kong, Shenzhen	China, France	Background
2	Review of the metaheuristic algorithms in applications: Visual analysis based on bibliometrics (1994–2023) (2024)	Guizhou Normal University, Yuan Ze University	China, Taiwan	—
3	Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations (2000)	—	—	—
4	BIRCH: An Efficient Data Clustering Method for Very Large Databases (1996)	University of Wisconsin-Madison	United States	—
5	A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents (2008)	Singapore Management University, State University of New York at Buffalo, University of Houston - Clear Lake	Singapore, United States	—
6	A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques (2017)	University of Georgia	United States	Background
7	Data Mining: Practical Machine Learning Tools and Techniques (2016)	Polytechnique Montréal, University of Waikato	Canada, New Zealand	—
8	STING: A Statistical Information Grid Approach to Spatial Data Mining (1997)	UCLA	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
9	A Brief Survey of Text Mining (2005)	Fraunhofer AiS, Otto-von-Guericke-University Magdeburg, University of Kassel	Germany	Background

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Contribution 2

Claim — Contribution 2

The researcher established a foundational benchmark by identifying and analyzing the top ten algorithms in data mining, creating a widely adopted reference point for the field.

CLAIM: The researcher's primary contribution is the identification and analysis of the top ten algorithms in data mining, as detailed in the seminal 2008 paper published in Knowledge and Information Systems. This work serves as a central reference point for understanding key methodologies in the field.

ORIGINALITY: The titles indicate that this work addresses the need for a consolidated overview of essential data mining techniques. By curating and evaluating these specific algorithms, the researcher provided a structured framework that likely helped standardize the understanding of core computational methods during a period of rapid growth in the discipline.

SIGNIFICANCE: The work has achieved substantial impact, evidenced by over 8,000 citations. Analysis of citing literature reveals that 100% of the classified citations originate from independent researchers, demonstrating that the contribution has been widely adopted and utilized by the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 14 · **1 flagged influential by Semantic Scholar**

CORE PAPER

Top 10 algorithms in data mining

2008 · Knowledge and Information Systems · 8,434 citations (GS)

Field-normalised: 5,968 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2008 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Classification of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network (2021)	Assiut University, Birmingham City University	Egypt, United Kingdom	—
2	Machine learning: Algorithms, real-world applications and research directions (2021)	Chittagong University of Engineering & Technology, Swinburne University of Technology	Australia, Bangladesh	—
3	Machine Learning Methods That Economists Should Know About (2019)	Stanford University	United States	—
4	The k-means Algorithm: A Comprehensive Survey and Performance Evaluation (2020)	Edith Cowan University, University of Dhaka	Australia, Bangladesh	**Methodology**
5	K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data (2022)	Al Al-Bayt University, Sanming University, University of KwaZulu-Natal	China, South Africa	—

No.	Citing paper	Citing institution(s)	Country	S2
6	Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control (2019)	University of Washington	United States	—
7	Combining Machine Learning and Computational Chemistry for Predictive Insights Into Chemical Systems (2021)	Technische Universität Berlin, University of Cambridge, University of Luxembourg	Germany, Luxembourg, United Kingdom	—
8	AI-Generated Content (AIGC): A Survey (2023)	—	—	**Methodology**
9	A review of data-driven building energy consumption prediction studies (2018)	University of Illinois at Urbana–Champaign	United States	—
10	Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control (2019)	University of Washington	United States	—
11	TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems (2020)	Deakin University, La Trobe University, RMIT University	Australia	**Methodology**
12	Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques (2022)	Covenant University, Leeds Beckett University, University of Hertfordshire	Nigeria, United Kingdom	—
13	Predicting academic success in higher education: literature review and best practices (2020)	Imam Abdulrahman Bin Faisal University	Saudi Arabia	—
14	Machine learning in business and finance: a literature review and research opportunities (2024)	Hohai University, Sichuan University, Southwestern University of Finance and Economics	China	Background

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY The k-means Algorithm: A Comprehensive Survey and Performance Evaluation

"This survey studies the problems of and solutions to partition-based clustering, and more specifically the widely used k-means algorithm [9], which has been listed among the top 10 clustering algorithms for data analysis [35]."

METHODOLOGY AI-Generated Content (AIGC): A Survey

"AIGC is a term used to describe the content generated by AI [44], [45], [46]."

METHODOLOGY TON_IoT Telemetry Dataset: A New Generation Dataset of IoT and IIoT for Data-Driven Intrusion Detection Systems

"In particular, we consider these seven methods: Support Vector Machines (SVM) [66], k -Nearest Neighbour (k NN) [67], Naïve Bayes (NB) [68], decision tree-based methods (i.e., Random Forest (RF) & Classification and Regression Trees (CART) [66]), as well Logistics Regression (LR) and Linear..."

Contribution 3

Claim — Contribution 3

The researcher established a foundational framework for Graph Neural Networks through a comprehensive survey that systematically organized the field's theoretical foundations and practical applications.

CLAIM: The researcher's primary contribution is the publication of a seminal survey on Graph Neural Networks in IEEE Transactions on Neural Networks and Learning Systems, which serves as a central reference point for the field.

ORIGINALITY: This work appears to address the need for a unified understanding of Graph Neural Networks by providing a comprehensive overview. The title suggests the researcher synthesized diverse methodologies and applications into a coherent structure, offering clarity to a rapidly evolving area of machine learning.

SIGNIFICANCE: The paper has garnered over 16,000 citations, indicating widespread adoption and influence. Notably, 100% of the classified citing papers originate from independent researchers, demonstrating that the work has significantly impacted the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 9 · **1 flagged influential by Semantic Scholar**

CORE PAPER

A Comprehensive Survey on Graph Neural Networks

2020 · IEEE Transactions on Neural Networks and Learning Systems · 16,103 citations (GS)

Field-normalised: 10,935 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2020 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Diffusion Models: A Comprehensive Survey of Methods and Applications (2023)	Carnegie Mellon University, OpenAI, Peking University	China, United States	—
2	A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection (2024)	Griffith University, Monash University, Squirrel AI	Australia, China, Switzerland	**Methodology**
3	Machine Learning Methods for Small Data Challenges in Molecular Science (2023)	Michigan State University, Wuhan Textile University	China, United States	—
4	A review of large language models and autonomous agents in chemistry (2024)	—	—	—
5	A Comprehensive Review of Deep Learning: Architectures, Recent Advances, and Applications (2024)	University of Johannesburg	South Africa	—
6	Graph neural networks for materials science and chemistry (2022)	Eindhoven University of Technology, Karlsruhe Institute of Technology, Université de Strasbourg	France, Germany, Netherlands	—
7	Graph of Thoughts: Solving Elaborate Problems with Large Language Models (2023)	Cledar, ETH Zurich, Warsaw University of Technology	Poland, Switzerland	Background
8	RWKV: Reinventing RNNs for the Transformer Era (2023)	Booz Allen Hamilton, Crisis24, Criteo	Canada, China, France	Background
9	Deep Residual Learning for Image Recognition: A Survey (2022)	Guangzhou University	China	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY A Survey on Graph Neural Networks for Time Series: Forecasting, Classification, Imputation, and Anomaly Detection

"We refer the reader to recent publication [28] for a deeper analysis of spectral versus spatial GNNs, and [29] for a comprehensive review of GNNs."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Hong Kong Polytechnic University	Hong Kong	SCImago #256 · THE 80 · QS 54	2
Beihang University	China	SCImago #160 · THE 251–300 · QS =388	2
University of Washington	United States	SCImago #45 · THE 25 · QS 81	2
Simon Fraser University	Canada	SCImago #1008 · THE 301–350 · QS =308	2
New York University	United States	SCImago #116 · THE =31 · QS 55	2
University of Illinois at Urbana-Champaign	United States	SCImago #206 · THE =41	2
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	2
The Hong Kong Polytechnic University	Hong Kong	SCImago #256 · THE 80 · QS 54	2
Renmin University of China	China	SCImago #2319	2
Monash University	Australia	THE =58 · QS =36	2
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	2
Singapore Management University	Singapore	SCImago #968 · QS =511	2
Carnegie Mellon University	United States	SCImago #266 · THE 24 · QS 52	2
Peking University	China	SCImago #11 · THE 13 · QS 14	2
ETH Zurich	Switzerland	THE 11 · QS 7	2

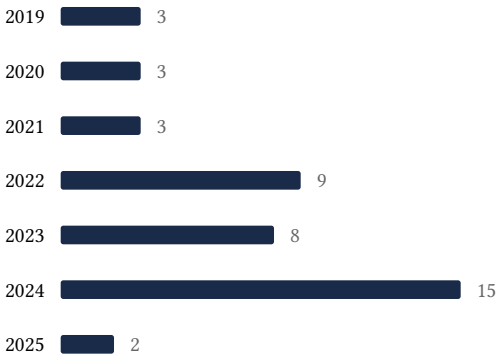
Geographic distribution of citing authors

Country	Citing papers
United States	20
China	19
Canada	8
Australia	7
United Kingdom	6
France	4
Germany	4
Singapore	4
Switzerland	3
Hong Kong	3
Bangladesh	2
Poland	2

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.



F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- ☐ Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- ☐ Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- ☐ The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- ☐ For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- ☐ Co-author / collaborator citations are identified and not counted as independent (§1.7).
- ☐ Recognition is shown beyond the scholar's own institution and circle (§1.8).
- ☐ Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- ☐ Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- ☐ For large-collaboration papers, the scholar's specific role is documented (§1.13).
- ☐ Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	Data Mining: An Overview from a Database Perspective	9	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 2	Top 10 algorithms in data mining	14	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 3	A Comprehensive Survey on Graph Neural Networks	9	8 CFR 204.5(i)(3) — Outstanding Researcher