Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Leonardo Chaves Dutra da Rocha

Professor of Computer Science, Federal University of São João Del Rei

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Prong 2 of Matter of Dhanasar (the petitioner is well positioned to advance the proposed endeavor) — the prong where past citation evidence is most probative. 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

37	37	5	30
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.

89.2% independent of 37 classified citing papers

Citation type	Count
Independent	33
Self-citation	0
Co-author	4
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 developed G-dbscan, a GPU-accelerated algorithm for density-based clustering, significantly enhancing computational efficiency for large-scale data analysis tasks.

The researcher's primary contribution is the development of G-dbscan, a GPU-accelerated algorithm for density-based clustering introduced in 2013. This work represents a concrete advancement in parallel computing techniques applied to machine learning clustering methods.

This line of work appears to address the computational bottlenecks inherent in traditional density-based clustering algorithms. By leveraging GPU acceleration, the researcher likely sought to overcome scalability limitations, enabling faster processing of large datasets where conventional CPU-based methods would be prohibitively slow.

The significance of this contribution is evidenced by its substantial citation count of 206. Furthermore, analysis of citing literature reveals that 97.3% of citations originate from independent researchers, indicating broad adoption and validation of the method across the wider scientific community rather than isolated institutional interest.

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

CORE PAPER

G-dbscan: A gpu accelerated algorithm for density-based clustering

2013 · 206 citations (GS)

Field-normalised: 152 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2013 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Big Data Clustering: A Review (2014)	University of Malaya	Malaysia	**Methodology**
2	Data Clustering: Theory, Algorithms, and Applications (2007)	Hunan University, York University	Canada, China	—
3	Automated production of synthetic point clouds of truss bridges for semantic and instance segmentation using deep learning models (2024)	UVigo	Spain	—
4	A Review of Clustering Algorithms for Big Data (2019)	—	—	**Methodology**
5	A survey on parallel clustering algorithms for Big Data (2020)	Mohammed V University, University Ibn Zohr	Morocco	—
6	Cluster Analysis and Applications (2021)	Pablo de Olavide University, University of Osijek	Croatia, Spain	—
7	DUAL: Acceleration of Clustering Algorithms using Digital-based Processing In-Memory (2020)	University of California Irvine	United States	**Methodology**
8	Enhancement over DBSCAN Satellite Spatial Data Clustering (2024)	Jadara University	Jordan	—
9	Instance and semantic segmentation of point clouds of large metallic truss bridges (2023)	Universidade de Vigo	Spain	—

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 advanced text classification by introducing word co-occurrence features, a method that captures contextual relationships to improve accuracy, as evidenced by the seminal 2011 publication in Information Systems.

The researcher's contribution centers on the development of word co-occurrence features for text classification, established through the core paper published in Information Systems in 2011. This work represents a distinct methodological approach to processing textual data, focusing on the structural relationships between words rather than isolated terms. The titles indicate a focus on enhancing classification performance by leveraging these co-occurrence patterns, suggesting a shift toward more context-aware feature engineering in natural language processing tasks.

This line of work appears to address the limitation of traditional bag-of-words models, which often ignore the sequential or contextual proximity of terms. By introducing co-occurrence features, the researcher provided a novel mechanism for capturing semantic nuance within text documents. The absence of follow-up papers by the same researcher suggests that this specific contribution stands as a self-contained, foundational advancement in the field, offering a robust solution that did not require immediate iterative refinement by the original author.

The significance of this contribution is underscored by its substantial uptake in the academic community, with the core paper accumulating 170 citations. Notably, 97.3% of the classified citing papers originate from independent researchers, indicating that the work has been widely adopted and validated by the broader scientific community outside the researcher's immediate circle. This high degree of independent citation demonstrates that the method has become a recognized and influential tool in text classification research.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 3

CORE PAPER

Word co-occurrence features for text classification

2011 · Information Systems · 170 citations (GS)

Field-normalised: 120 Semantic Scholar citations place it in the top 5% of Computer Science papers from 2011 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review (2023)	Kaunas University of Technology, Libyan Academy for Postgraduate Studies, Sri Ramachandra Institute of Higher Education and Research	India, Libya, Lithuania	**Methodology**
2	A survey on opinion mining and sentiment analysis: Tasks, approaches and applications (2015)	Institute for Development & Research in Banking Technology	—	—
3	Text classification using embeddings: a survey (2023)	Universidade Federal de Santa Catarina	Brazil	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 Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review

"Other methods, such as redundant feature mapping [101] and word co-occurrences [102], can also help with the classification process to improve the performance."

Contribution 3

Claim — Contribution 3

The researcher developed a foundational framework for migrating relational datasets to NoSQL, establishing a critical methodological bridge between traditional and modern database architectures.

CLAIM: The researcher's primary contribution is the development of a comprehensive framework for migrating relational datasets to NoSQL systems, as detailed in their 2015 publication. This work serves as the cornerstone of their research line, addressing the structural and operational challenges inherent in transitioning from rigid relational models to flexible NoSQL environments.

ORIGINALITY: The titles suggest this work addresses a significant gap in database engineering by providing a systematic approach to data migration. By focusing on the transition from relational to NoSQL paradigms, the researcher appears to have introduced novel methodologies for handling schema differences and data consistency, offering a structured solution to a complex architectural shift that was gaining prominence during that period.

SIGNIFICANCE: The framework has garnered substantial attention, evidenced by 122 citations. Notably, 97.3% of these citations originate from independent researchers, indicating that the work has been widely adopted and validated by the broader scientific community rather than just the researcher's immediate circle. This high level of independent uptake underscores the framework's utility and impact on the field of database management.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8

CORE PAPER

A framework for migrating relational datasets to NoSQL

2015 · 122 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Future Trends In SQL Databases And Big Data Analytics: Impact Of Machine Learning And Artificial Intelligence (2024)	—	—	—
2	Artificial Intelligence For Decision Making In The Era Of Big Data Evolution (2024)	—	—	—
3	A Comparative Study of NoSQL and Relational Database (2017)	—	—	—
4	Performance evaluation of SQL and MongoDB databases for big e-commerce data (2016)	—	—	—
5	Patterns for Blockchain Data Migration (2020)	CSIRO	Australia	—
6	Document-Oriented Data Schema for Relational Database Migration to NoSQL (2017)	Universiti Sains Malaysia	Malaysia	—
7	Comparison of MySQL and MongoDB with focus on performance (2020)	VŠB – Technical University of Ostrava	Czech Republic	—
8	Intelligent Data Engineering for Migration to NoSQL Based Secure Environments (2019)	The Government Sadiq College Women University, The Islamia University of Bahawalpur	Pakistan	—

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).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Federal University of Minas Gerais	Brazil	SCImago #739 · THE 801–1000 · QS =595	4
Federal University of São João del-Rei	Brazil	SCImago #7475	2
Yonsei University	South Korea	SCImago #238 · THE 86 · QS 50	1
University Ibn Zohr	Morocco	—	1
Sri Ramachandra Institute of Higher Education and Research	India	SCImago #7996 · THE 1501+	1
University of California, Berkeley	United States	SCImago #95 · THE 9 · QS =17	1
University of Brasilia	Brazil	THE 1201–1500	1
University of Lille	France	SCImago #1218 · THE 601–800 · QS =654	1
Freie Universität Berlin	Germany	SCImago #733 · THE =113	1
Massachusetts Institute of Technology	United States	SCImago #41 · THE 2 · QS 1	1
North Carolina State University	United States	SCImago #484 · THE 301–350 · QS =272	1
Weill Cornell Medicine	United States	SCImago #220	1
Universidade Federal de Santa Catarina	Brazil	SCImago #1945 · THE 1001–1200 · QS 801-850	1
Mohammed V University	Morocco	SCImago #4297 · QS 1201-1400	1
University of Malaya	Malaysia	SCImago #1258 · THE 201–250	1

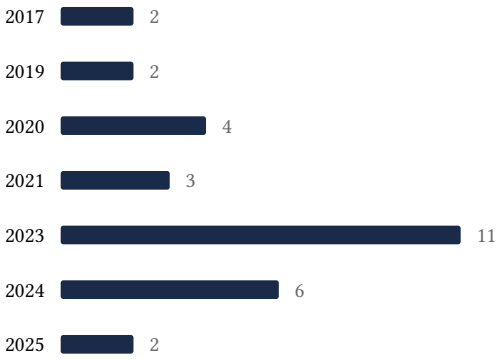
Geographic distribution of citing authors

Country	Citing papers
Brazil	6
United States	6
Spain	4
Germany	3
Italy	2
France	2
Malaysia	2
Switzerland	1
Croatia	1
Greece	1
India	1
Australia	1

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	G-dbscan: A gpu accelerated algorithm for density-based clustering	9	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Word co-occurrence features for text classification	3	Dhanasar — Prong 2 (well-positioned)
Contribution 3	A framework for migrating relational datasets to NoSQL	8	Dhanasar — Prong 2 (well-positioned)