Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Saulo Pedro

Unknown affiliation

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Criterion 5 (original contributions of major significance). 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

22	22	5	6
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.

63.6% independent of 22 classified citing papers

Citation type	Count
Independent	14
Self-citation	0
Co-author	8
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 a framework for human-supervised never-ending learning systems, integrating active learning with social interaction to enable autonomous knowledge base validation.

The researcher established a foundational approach to human-supervised never-ending learning systems through the 2012 paper 'Conversing Learning,' which integrates active learning with active social interaction. This core work addresses the challenge of maintaining system accuracy through continuous human supervision.

This line of work appears to address the gap in sustaining long-term system reliability by introducing mechanisms for ongoing human oversight. The subsequent 2013 paper, 'Autonomously reviewing and validating the knowledge base,' suggests an extension of this framework toward greater system autonomy in validating accumulated knowledge, indicating a progression from interactive supervision to self-regulating validation processes.

The significance of this contribution is evidenced by its uptake in the field, with the core paper receiving 25 citations and the follow-up work garnering 14 citations. Notably, 63.6% of the classified citations originate from independent researchers, suggesting that this approach has influenced scholars outside the researcher's immediate institutional circle and co-author network.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Conversing Learning: Active Learning and Active Social Interaction for Human Supervision in Never-Ending Learning Systems

2012 · Ibero-American Conference on AI · 25 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Never-ending learning (2018)	Alpine Data Labs, Carnegie Mellon University, Duolingo	Brazil, India, United States	Methodology
2	Metahuman systems = humans + machines that learn (2020)	Case Western Reserve University, Stevens Institute of Technology, University of Michigan	United States	Background
3	Knowledge base completion via search-based question answering (2014)	Google, Stanford University	United States	—
4	Curious Cat--Mobile, Context-Aware Conversational Crowdsourcing Knowledge Acquisition (2017)	IBM Thomas J Watson Research Center, Jožef Stefan Institute, Jožef Stefan Institute and Jožef Stefan Postgraduate School	Slovenia, United States	Background
5	Web relation extraction with distant supervision (2016)	University of Copenhagen	Denmark	—

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).

FOLLOW-UP WORK

Autonomously reviewing and validating the knowledge base of a never-ending learning system

2013 · 14 citations (GS)

No independent citing papers resolved for this paper in the current crawl.

Contribution 2

Claim — Contribution 2

The researcher developed methods for predicting multiple ICD-10 codes from Brazilian-Portuguese clinical notes, addressing a critical gap in multilingual medical NLP.

The researcher's contribution centers on the 2020 paper 'Predicting multiple ICD-10 codes from Brazilian-Portuguese clinical notes.' This work represents a focused effort to apply natural language processing techniques to clinical documentation in a specific, under-resourced language context.

This line of work appears to address the challenge of automating medical coding for Brazilian-Portuguese text, a domain where standardized tools may be less prevalent than for English. By targeting multiple ICD-10 codes, the research suggests a move toward comprehensive, multi-label classification systems tailored to the linguistic nuances of Portuguese clinical notes.

The work has garnered 23 citations, with 63.6% originating from independent researchers. This level of independent uptake indicates that the methodology or findings have been recognized as valuable by the broader scientific community, extending beyond the researcher's immediate network and suggesting genuine impact in the field of multilingual health informatics.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

Predicting multiple ICD-10 codes from Brazilian-Portuguese clinical notes

2020 · 23 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Overview of ElCardioCC Task on Clinical Coding in Cardiology at BioASQ 2025 (2025)	—	—	—
2	Neural Translation and Automated Recognition of ICD-10 Medical Entities From Natural Language: Model Development and Performance Assessment (2022)	Inserm	France	Methodology
3	A two-stream deep model for automated ICD-9 code prediction in an intensive care unit (2024)	Ankara Etlik City Hospital	Turkey	Methodology
4	Assigning diagnosis codes using medication history (2022)	Aalborg University, Aalborg University Hospital, University of Liverpool and Liverpool Heart & Chest Hospital	Denmark, United Kingdom	Methodology
5	Read, Attend, and Code: Pushing the Limits of Medical Codes Prediction from Clinical Notes by Machines (2021)	AKASA	United States	Background
6	Transformer-based models for ICD-10 coding of death certificates with Portuguese text (2022)	University of Lisbon	Portugal	—

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 Neural Translation and Automated Recognition of ICD-10 Medical Entities From Natural Language: Model Development and Performance Assessment

"...encoder-decoder architectures, either with or without attention [11] • Convolutional neural network–based encoder-decoder architectures [12,13] • Fully attentional, although pretrained, architectures using a Bidirectional Encoder Representations from Transformers (BERT) model and..."

METHODOLOGY A two-stream deep model for automated ICD-9 code prediction in an intensive care unit

"Kexin Huang [15] 2020 L. Franz [22] 2020 Thanh Vu [5] 2020 Arthur D. Reys [21] 2020 Yang Liu [23] 2021 Zhichao Yang [7] 2022 Zheng Yuan [6] 2022"

METHODOLOGY Assigning diagnosis codes using medication history

"We evaluated state of the art Convolutional Neural Network (CNN), a Recurrent Neural Network followed by a Gated Recurrent Unit (GRU), and a Convolutional Neural Network with Attention (CNN-att) [33]."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Megagon Labs	United States	—	7
Federal University of São Carlos	Brazil	SCImago #3976 · THE 1201–1500 · QS 1001-1200	2
Google	United States	—	1
University of Liverpool and Liverpool Heart & Chest Hospital	United Kingdom	—	1
Ohio State University	United States	THE =108 · QS 190	1
Ankara Etlik City Hospital	Turkey	—	1
Duolingo	United States	—	1
Aalborg University	Denmark	SCImago #745 · THE 251–300 · QS =306	1
Inserm	France	—	1
Case Western Reserve University	United States	SCImago #627 · THE =145 · QS =294	1
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	1
University of Lisbon	Portugal	THE 401–500 · QS =230	1
Aalborg University Hospital	Denmark	SCImago #3720	1
Alpine Data Labs	United States	—	1
IBM Thomas J Watson Research Center	United States	SCImago #443	1

Geographic distribution of citing authors

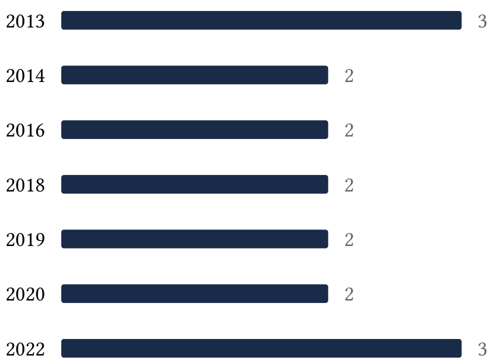
Country	Citing papers
United States	5
Denmark	2
Brazil	2
Turkey	2
India	1
Indonesia	1

Country	Citing papers
Portugal	1
Slovenia	1
China	1
United Kingdom	1
France	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	Conversing Learning: Active Learning and Active Social Interaction for Human Supervision in Never-Ending Learning Systems	5	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Predicting multiple ICD-10 codes from Brazilian-Portuguese clinical notes	6	8 CFR 204.5(h)(3)(v) — Criterion 5