

Teng-Jui Lin

1 May 2023

CHEM E 499

**UW Library Research Award for Undergraduates Research Strategies Reflective Essay:
Library Resources Aid Efficient Interdisciplinary Research**

Saving and bettering people's lives has always driven me to explore the frontier of nanomedicine delivery. As an interdisciplinary chemical engineer, I am also passionate about mathematically modeling complex biomedical systems like the brain and analyzing large volumes of high-dimensional data. In my two years in the disease-directed engineering lab supervised by Professor Nance in the Department of Chemical Engineering, I combined my interests in nanomedicine, neuroscience, and data science to solve an interdisciplinary problem: How can we quantify cell shape's response to diseases using machine learning, take advantage of the growing tools in data science, and use the results to develop therapeutics?

To formulate a research question taking advantage of the lab's expertise on neuroinflammatory diseases, I searched Google Scholar¹ with broad terms such as "microglia" (the immune cells of the brain) and "cell morphology" (shape) separately with the filter "review article" to gain a comprehensive understanding of the cell type of interest and the shape analysis challenges for all cells. Although most articles are informative, some are irrelevant to my research direction. I realized that Google Scholar is indexing many articles not specific to biomedicine. After consulting with the UW Health Science Library (HSL) website², I conducted further searches in PubMed³ instead to ensure the relevance of the articles to my research.

I dove deeper into the intersection of microglia and cell shapes by using the AND operator to limit the scope and identify methods used at the research frontier. From the referenced research articles in the review articles, I gained a preliminary understanding of the field standard in evaluating microglia shapes. Because the referenced research articles are dated, I conducted another search on microglia morphology by limiting the earliest published time to "2018", which provided insights into how the field evolves in research methods (from statistical analysis to machine learning) and interpretations of conflicting findings (from microglia's binary pro- or anti-inflammatory states to a spectrum of states that contributes differently to disease progression). I further conducted similar searches on preprint servers (arXiv⁴ and bioRxiv⁵) to identify recent studies that have not yet been through the peer review process, giving me a standard for state-of-the-art research. The search process took many iterations of changing keywords and filters, especially when I found vocabulary words and conflicting results that required further research.

During the search process, I needed clear inclusion criteria for citation and in-depth article reading. To set such criteria, I adopted the CRAAP test⁶ suggested by the UW Libraries Research Guides, especially for identifying recent and relevant articles for my research. Because of the interdisciplinary nature of my research, I used articles describing one single discipline for background research and placed emphasis on interdisciplinary articles that focus on microglia morphology specifically for methods research. To ensure the credibility of the articles, I prioritized articles published in peer-reviewed journals compared to preprint servers. Although the prestige and impact factor of journals does not necessarily imply better article quality, I prioritized more established journals in the field (*Bioengineering & Translational Medicine* and *Nature Neuroscience*) as a growing researcher before I can independently judge the merit of research studies.

With all the articles on hand, managing citations and finding the connections between articles became challenging. To manage the articles and reading notes more efficiently, I adopted Zotero⁷, an open-source citation manager, and discovered useful tagging and classification features beyond automatically generating citations in Word. The browser extension of Zotero made article citation a breeze, greatly speeding up my article identification process. With >1000 articles in my library, Zotero has become an integral part of my learning and academic writing process beyond research. Although I can manage all the articles easily, the connection between them is not obvious: the authors can only cite references but cannot backtrack articles citing themselves. To overcome such issue, I used Semantic Scholar⁸ to track citations of articles of interest. Such feature is handy for identifying research studies that have been done based on key methods/protocol papers. To identify connections between research papers, I used Connected Papers⁹, a citation visualization tool that identifies major articles related to the article of interest, to gain a bigger picture of the research area, whether in terms of key results and methods or prominent research groups in the field.

As I gained a comprehensive view of my research area of microglia morphology, I noticed that quantitative methods for microglia microscopy images, especially those with image processing, are lacking. With my data science background, I realized such gut feeling of disparity can be quantified by the number of articles indexed on PubMed. I first standardized the search terminology using Medical Subject Headings (MeSH)¹⁰ recommended by UW HSL². I then searched microglia, confocal microscopy, and image processing individually and combinatorically with AND operator with and without MeSH standardization for each year after 1990, recording the number of indexed publications on PubMed. I found that the number of microglia microscopy publications with image processing is only 1/10 of the total number of

microglia microscopy publications, suggesting a lack of automated image processing methods for quantifying microglia microscopy images for shape analysis. The number of publications indexed with MeSH terms is only half the number of publications indexed without MeSH terms, suggesting a lack of standardization of the field at article tagging, which hinders article accessibility. Because the search process is labor intensive, I automated the search process with web scrapping packages (Requests¹¹ and BeautifulSoup4¹²) in Python, being 93% more efficient than manual search. The results are generalizable to many cell types in the brain, found with a simple tweak of search keywords. In collaboration with my mentor, the generalized results are documented in a submitted review article¹³ detailing the current practices of image processing for cell images.

As the project concludes, I have become a researcher well-versed in literature search, interpretation, citation, and data mining. With the resources provided by UW Libraries, I synthesized my interdisciplinary domain knowledge with the state-of-the-art method to effectively address my research question, further developing my critical thinking skills.

References

1. Google Scholar. Accessed May 1, 2023. <https://scholar.google.com/>
2. Maloy C. Library Guides: PubMed at the UW: Home. Accessed May 1, 2023. <https://guides.lib.uw.edu/hsl/pubmed/home>
3. PubMed. PubMed. Accessed May 1, 2023. <https://pubmed.ncbi.nlm.nih.gov/>
4. arXiv.org. Accessed May 1, 2023. <https://arxiv.org/>
5. bioRxiv.org. Accessed May 1, 2023. <https://www.biorxiv.org/>
6. Tawatao C. Library Guides: FAQ: How do I know if my sources are credible/reliable? Accessed May 1, 2023. <https://guides.lib.uw.edu/research/faq/reliable>
7. Zotero. Accessed May 1, 2023. <https://www.zotero.org/>
8. Semantic Scholar. Accessed May 1, 2023. <https://www.semanticscholar.org/>
9. Connected Papers. Accessed May 1, 2023. <https://www.connectedpapers.com/>
10. PubMed User Guide. PubMed. Accessed May 1, 2023. <https://pubmed.ncbi.nlm.nih.gov/help/>
11. Requests 2.29.0 documentation. Accessed May 1, 2023. <https://requests.readthedocs.io/en/latest/>
12. Beautiful Soup 4.12.0 documentation. Accessed May 1, 2023. <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>
13. Helmbrecht H, Lin TJ, Janakiraman S, Decker K, Nance E. Prevalence and Practices of Immunofluorescent Cell Image Processing: a Systematic Review. *Submitted*.