

Applying the Science of Systematic Review and Meta-Analysis to Retrospective Artificial Intelligence Based Studies: The importance of performance evaluation

MASSOUD SOKOUTI¹, BABAK SOKOUTI²

¹Nuclear Medicine Research Center, Mashhad University of Medical Sciences, Mashhad, Iran

²Biotechnology Research Center, Tabriz University of Medical Sciences, Tabriz, Iran

Correspondence to Dr. Babak Sokouti, Email: b.sokouti@gmail.com; sokoutib@tbzmed.ac.ir, Tel: +98 41 33364038 Fax: +98 41 33370420

The rationale behind the meta-analysis goes back to the 17th century studies of astronomy which then Karl Pearson performed a study based on meta-analysis using the data for typhoid inoculation in 1904¹. After, William Cochran applied this type of analysis to medical researches by taking the advantage of multiple previous studies^{2,3}. For more information and details on the history, the readers are referred to⁴.

To emerge the important role of systematic and meta-analysis studies even in the area of artificial intelligence systems, it is anticipated that more reliable results can be driven from previous research studies alongside a simple review of such studies from which most of them may be ignored or not included as a matter of their non-systematical type of reviews. The meta-analysis technique uses various types of statistics tools and methodologies to commonly derive a predictive diagnostic or non-diagnostic performance result of their compared corresponding approaches on the target defined disorders using information included in different datasets of previous studies. Although, a meta-analysis study can be regarded as a review of previous studies, however, it thoroughly targets not only the achieving results of those studies but also determine the in-common and non-common patterns of those researches as well as biases of the performance results whether they have been inserted intentionally or unintentionally. The importance of meta-analysis has been vastly discussed in medical sciences and therefore, been conducted rigorously through various studies, mostly on clinical trial ones. However, this technique is one of those less valued tools imported in to biomedical engineering studies and hence, their related algorithms mostly on the performance of artificial intelligence approaches. One of those studies to mention is the one performed on classification algorithms for pattern recognition by So Young Sohn in 1999 based on some in-house implemented statistics tools without considering the meta-analysis software⁵. Moreover, in 2015, Horn et al have conducted a systematic review on functional brain imaging studies on assessing the familiarity of artificial neural networks and discussed their pros and cons in terms of their experimental conflicting results based on a meta-analysis on 68 published articles⁶. In another recent study, the role of real-time biomedical systems has been evaluated by a meta-analysis approach on 134 real-time papers in terms of computational complexity, delay and speed up considering various types of algorithms and hardware implementation⁷. Recently, two types of systematic review and analysis have been performed which shows the potential non-mature trends of this approach in artificial

intelligence based researches. In the first one the authors studied the performance of different machine learning algorithms for heart disease diagnosis⁸; however, the meta-analysis part was not performed due to the existence of heterogeneity in the final included studies through the PRISMA (Preferred reporting items for systematic reviews and meta-analyses) checklist^{9,10}. And in the second one, the performance of several DNA based encryption algorithms based according to the results obtained from previous publications has been proposed where, it has been found out that there were no improvements in the proposed algorithms and it has been suggested that a dataset of images should be available in order to test and evaluate the performance of methodologies. However, the methodologies should also be available for public use¹¹. Moreover, the analyses section can be carried out through a simple statistical student's t test analysis or the meta-analysis procedure using available tools such as Meta-Disc¹², MIX¹³, and Meta-Analyst¹⁴.

While comparing the two environments (i.e., clinical and computational), there are in-common units for decision making in diagnosing symptoms which are human (brain system and some data) and computer (artificial intelligence systems and some data). This outstanding feature and the abovementioned examples makes the meta-analysis studies applicable to the researches performed based on artificial intelligence systems, too. This will open a new view on interactions between the results obtained from previous studies while considering their special algorithms, different datasets, and possible biases.

One more thing to emphasize for the future research studies is on publicizing the datasets and the implemented algorithms in terms of web servers, Java, C++ and Matlab libraries or R packages to make the results re-generable using new datasets which make them more comparable with new designed methodologies to ease the meta-analysis robust studies. As, it is also clear, most of the web servers and datasets in the medical parts coupled with data derived from bioscience knowledge are publicly¹⁵⁻²⁰.

REFERENCES

1. Pearson K. Report on Certain Enteric Fever Inoculation Statistics. *British medical journal*. 1904;2(2288):1243-6.
2. Cochran WG. Problems arising in the analysis of a series of similar experiments. *J Roy Stat Soc*. 1937;4:102-18
3. Yates F, Cochran WG. The analysis of groups of experiments. *J Agric Sci*. 1938;28:556-80.
4. O'Rourke K. An historical perspective on meta-analysis: dealing quantitatively with varying study results. *Journal of the Royal Society of Medicine*. 2007;100(12):579-82.

5. Sohn SY. Meta Analysis of Classification Algorithms for Pattern Recognition. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*. 1999;21(11):1137-44.
6. Horn M, Jardri R, D'Hondt F, Vaiva G, Thomas P, Pins D. The multiple neural networks of familiarity: A meta-analysis of functional imaging studies. *Cognitive, affective & behavioral neuroscience*. 2015.
7. Faust O, Yu W, Rajendra Acharya U. The role of real-time in biomedical science: a meta-analysis on computational complexity, delay and speedup. *Computers in biology and medicine*. 2015;58:73-84.
8. Safdar S, Zafar S, Zafar N, Khan NF. Machine learning based decision support systems (DSS) for heart disease diagnosis: a review. *Artificial Intelligence Review*. 2017.
9. Knobloch K, Yoon U, Vogt PM. Preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement and publication bias. *J Craniomaxillofac Surg*. 2011;39(2):91-2.
10. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097.
11. Sokouti M, Sokouti B. A PRISMA-compliant systematic review and analysis on color image encryption using DNA properties. *Computer Science Review*. 2018;29:14-20.
12. Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31.
13. Bax L, Yu LM, Ikeda N, Tsuruta H, Moons KG. Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC Med Res Methodol*. 2006;6:50.
14. Wallace BC, Schmid CH, Lau J, Trikalinos TA. Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data. *BMC Medical Research Methodology*. 2009;9(1):80.
15. Chen X, Yan CC, Zhang X, Zhang X, Dai F, Yin J, et al. Drug-target interaction prediction: databases, web servers and computational models. *Briefings in bioinformatics*. 2015.
16. Sokouti B, Rezvan F, Yachdav G, Dastmalchi S. GPCRTOP: A novel g protein-coupled receptor topology prediction method based on hidden Markov model approach using Viterbi algorithm. *Current Bioinformatics*. 2014;9(4):442-51.
17. Sokouti B, Rezvan F, Dastmalchi S. GPCRTOP v.1.0: One-step web server for both predicting helical transmembrane segments and identifying G protein-coupled receptors. *Current Bioinformatics*. 2017;12(1):80-4.
18. Tuncbag N, Kar G, Keskin O, Gursoy A, Nussinov R. A survey of available tools and web servers for analysis of protein-protein interactions and interfaces. *Briefings in bioinformatics*. 2009;10(3):217-32.
19. Unni S, Huang Y, Hanson RM, Tobias M, Krishnan S, Li WW, et al. Web servers and services for electrostatics calculations with APBS and PDB2PQR. *Journal of computational chemistry*. 2011;32(7):1488-91.
20. Sokouti B, Rezvan F, Dastmalchi S. Applying random forest and subtractive fuzzy c-means clustering techniques for the development of a novel G protein-coupled receptor discrimination method using pseudo amino acid compositions. *Molecular bioSystems*. 2015;11(8):2364-72.