**Research Paper**

Automatically Appraising the Credibility of Vaccine-Related Web Pages Shared on Social Media: A Twitter Surveillance Study

**Five-section Paper Review (498 words)**

**Research Question**

To develop an automatic appraisal system measuring the credibility of vaccine-related web pages shared on Twitter, two research questions were proposed.

1. What is the potential reach of the vaccine-related web page embedded on tweets?
2. How to measure and characterize the sharing and potential reach of those posts?

**Research Significance**

1. *Harmful Misinformation*

The spread of health-related misinformation detrimentally affects the public cognition of health issues, especially for the gullible individuals.

1. *Unqualified Criterion*

“Veracity” employed by previous researchers is inadequate for misinformation identification in the health domain.

1. *Inefficient Assessing Tools*

The existing tools used for credibility appraisal are inefficient and require expertise in health areas, impeding the progress of misinformation detection and elimination.

In contrast, this paper utilizes “credibility” as criterion and tailors a credibility checklist to measure the quality of vaccination-related web pages shared on Twitter.

**Research Method**

*Data Acquisition:*

Approximately 6 million English vaccine-related tweets and retweets between January 2017 and March 2018 were collected. After text preprocessing, 143,003 unique pages were selected for further analysis.

*Model Training:*

The training sample came from the other 474 web pages, whose credibility was manually evaluated by 3 investigators along with a 7-creteria-checklist appraisal tool. Applying Random Forest, Support Vector Machine, and Recurrent Neural Network to each criterion, 21 classifiers were developed, evaluated, selected, and assembled to an ensemble classifier, which calculates the credibility score by summing the total numbers of criteria each page satisfied.

*Result Analysis:*

With the ensemble identifier applied, those 143,003 pages were divided into low, median, and high credibility groups. Finally, potential exposure was estimated, and follower network was constructed.

**Research Finding**

The result proved that it is feasible to construct an automatic credibility appraisal of vaccine-related information, with 78% overall accuracy rate and over 96% at labeling low credibility pages. Compared with the medium- and high-credibility pages which opt to have a greater number of followers, tweets of low credibility (which accounted for the 11.86%) generated 9.34% of potential exposure. Overall, the majority of low-credibility web pages connected with small subpopulations while some had potential to reach millions of users. The proposed method could be used to identify potential communities at higher risk of exposure to low-credibility web pages.

**Validity Review**

Generally, the result is satisfactory and is consistent with previous studies, implying that the hypothesis validity existed. This paper made improvements by valuing the vaccine-related communication, by taking other critical factors (e.g. readability, clarity, transparency of sources) into consideration for health-related information assessment, and by calculating potential reach which ravel the sharing patterns of different subpopulation of credibility.

Limitations were elaborated, including potential problems caused by small training sample; failure to include other methods that might have better performance; absence of external validity of classifier (e.g. construct validity); limited characterization of vaccine-related web pages; and imperfect estimates of exposure. Specifically, limited characterization might result an underestimation of low-credibility communication shared on Twitter. Moreover, the missing of predefined search terms (such as injection) might also inhibit the semantic validity.

**Reference:**

Shah, Z., Surian, D., Dyda, A., Coiera, E., Mandl, K. D., & Dunn, A. G. (2019). Automatically Appraising the Credibility of Vaccine-Related Web Pages Shared on Social Media: A Twitter Surveillance Study. *Journal of medical Internet research*, *21*(11), e14007. https://doi.org/10.2196/14007