

## **Text mining in mosquito-borne disease: A systematic review**

### **ABSTRACT**

Mosquito-borne diseases are emerging and re-emerging across the globe, especially after the COVID19 pandemic. The recent advances in text mining in infectious diseases hold the potential of providing timely access to explicit and implicit associations among information in the text. In the past few years, the availability of online text data in the form of unstructured or semi-structured text with rich content of information from this domain enables many studies to provide solutions in this area, e.g., disease-related knowledge discovery, disease surveillance, early detection system, etc. However, a recent review of text mining in the domain of mosquito-borne disease was not available to the best of our knowledge. In this review, we survey the recent works in the text mining techniques used in combating mosquito-borne diseases. We highlight the corpus sources, technologies, applications, and the challenges faced by the studies, followed by the possible future directions that can be taken further in this domain. We present a bibliometric analysis of the 294 scientific articles that have been published in Scopus and PubMed in the domain of text mining in mosquito-borne diseases, from the year 2016 to 2021. The papers were further filtered and reviewed based on the techniques used to analyze the text related to mosquito-borne diseases. Based on the corpus of 158 selected articles, we found 27 of the articles were relevant and used text mining in mosquito-borne diseases. These articles covered the majority of Zika (38.70%), Dengue (32.26%), and Malaria (29.03%), with extremely low numbers or none of the other crucial mosquito-borne diseases like chikungunya, yellow fever, West Nile fever. Twitter was the dominant corpus resource to perform text mining in mosquito-borne diseases, followed by PubMed and LexisNexis databases. Sentiment analysis was the most popular technique of text mining to understand the discourse of the disease and followed by information extraction, which dependency relation and co-occurrence-based approach to extract relations and events. Surveillance was the main usage of most of the reviewed studies and followed by treatment, which focused on the drug-disease or symptom-disease association. The advance in text mining could improve the management of mosquito-borne diseases. However, the technique and application posed many limitations and challenges, including biases like user authentication and language, real-world implementation, etc. We discussed the future direction which can be useful to expand this area and domain. This review paper contributes mainly as a library for text mining in mosquito-borne diseases and could further explore the system for other neglected diseases.