AUTOMATIC EXTRACTION OF SPATIO-TEMPORAL MODELS OF WEST NILE VIRUS TRANSMISSION

Gabrielle A. Cooper, Oluwakemi O. Olawatuyi, Sean R. Scott, Toluola O. Oyeleye, Matthew N. Anyanwu, and Raphael D. Isokpehi

Center for Bioinformatics & Computational Biology, Department of Biology, Jackson State University, Jackson Mississippi 39217, USA

Abstract: Over the last decade, Human West Nile Virus (HWNV) infection has grown to be the dominant vector-borne disease in the United States. There is no specific treatment or vaccine to prevent the disease. Therefore, it is imperative to understand the environmental factors that favor the transmission of the virus from the mosquito to the bird, horse or human host. A challenge in devising effective West Nile Virus surveillance programs is the lack of sufficient empirical knowledge on the spatio-temporal transmission dynamics. The ever-growing number of publications on WNV infection makes it increasingly difficult to manually read all articles to identify empirical knowledge on transmission dynamics. Therefore, we developed an automatic approach to identify sentences from PubMed abstracts that contain descriptors of spatio-temporal models. We hypothesize that sentences in PubMed abstracts on West Nile Virus containing relevant descriptors will identify spatio-temporal models that have been used to understand transmission dynamics of West Nile virus. Our text mining approach extracted models of transmission dynamics for WNV the mosquito vector and hosts. For mosquito, models included raster-based mosquito abundance model, community level spatial models as well as mosquito abundance using human population density and land use/cover. Furthermore, a geographic information system (GIS) model relevant to mosquito was the geographic coverage and time for spraying mosquito adulticides. The collection of over 15,800 sentences from more than 2000 PubMed abstracts can be searched at http://compbio.jsums.edu/wnv

Keywords: Information Extraction, Spatio-Temporal Modeling, Text Mining, West Nile Virus

Acknowledgments: Mississippi NSF-EPSCoR “Modeling and Simulation of Complex Systems” Award (EPS-0903787); Pittsburgh Supercomputing Center's National Resource for Biomedical Supercomputing (T36 GM008789); U.S. Department of Homeland Security Science & Technology Directorate (2007-ST-104-000007; 2009-ST-062-000014; 2009-ST-104-000021); and Research Centers in Minority Institutions (RCMI) – Center for Environmental Health (NIH-NCRR G12RR13459). Disclaimer: “The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the funding agencies.