Regeneron WESEF 2023 Finalist

Noah Nager

Computational Biology and Bioinformatics

Implementation of Land Cover Data to Forecast West Nile Virus in the United States

West Nile virus (WNV) is the most impactful mosquito-borne disease in the United States, but is not currently treatable, meaning that patients are left to rely solely on supportive care. Since its arrival in the United States, it has caused over 2,500 deaths, which are compounded by the lasting impacts it leaves on the quality of life for those who survive infection. Because there are no available treatments, preventative measures are essential to limit the loss of life and quality of life due to WNV. An important form of preventative measure is the ability to predict and analyze the causes of the spread of disease, an effort that has primarily been undertaken using climatic models for WNV. In 2022, however, Nager developed a land cover-based model and argued in favor of it due to its superiority over climate-based models in both its advantage in long-range forecasting and its ability to provide actionable policy measures to prevent the spread of WNV. The 2022 model Nager developed solely analyzed the Gulf Coast region of the United States and was deterministic in nature. In this study, Nager's model was developed further by expanding the region of analysis to the entire contiguous United States and by reformulating the model into a probabilistic one. Six different neural networks that took in as inputs county land cover feature percentages and outputted Bernoulli-Gamma probability density functions for case incidences were trained as sub-models, in which the training data were split into regions of the United States, created by following the geographic distributions of the species of the various Culex genus mosquitos. The sub-models were found to be highly accurate, correctly capturing between 97% and 99% of test case incidences within a 90% prediction interval around the median. At a 10% prediction interval around the median, the sub-models were still accurate, correctly capturing between 65% and 75% of test case incidences. Additionally, when compared to the average county, counties with predicted elevated incidences of WNV had comparatively depressed pasture/hay, developed, deciduous forest, and shrub/scrub land cover proportions and comparatively elevated woody wetland, evergreen forest, grassland land cover proportions. Finally, the percentage change across the United States of the model's predicted median case incidence is a 14% increase from 2023 to 2100, with some regions like the Great Plains having much greater percentage increases. The analysis conducted in this study provides a novel method of forecasting the incidence of WNV in the United States and shows alterable factors that can be used to inhibit the spread of WNV. These results can be used as an early warning system to better inform policymakers in decisions regarding land cover changes, allowing for the reduction of the human cost of WNV.