



Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



Análise da influência de fatores ambientais nos casos de dengue registrados no ano de 2013 em Viçosa-MG utilizando algoritmos de aprendizado de máquina

Thales Silva Heck¹ – Departamento de Engenharia Civil – UFV; Júlio César de Oliveira² – Departamento de Engenharia Civil – UFV; Nilcilene das Graças Medeiros³ – Departamento de Engenharia Civil - UFV

Palavras-Chave: aprendizado de máquina, KNN, epidemiologia
Ciências Exatas e Tecnológicas – Geociências - Pesquisa

Introdução

Há muitos anos os epidemiologistas se beneficiam do geoprocessamento a fim de identificarem fatores ambientais (FA) em regiões propícias à proliferação de doenças e vetores específicos, e assim determinar suas possíveis causas.

Resultados e Discussão

Foram gerados modelos de regressão utilizando os algoritmos K-Nearest Neighbor (KNN), Random-Forest (RF), Bagging (BG) e Decision Tree (DT) e a Regressão linear multivariada (RLM). O algoritmo KNN apresentou $r^2 = 73,89\%$ e a RLM um $r^2 = 77,20\%$, sendo estes os melhores resultados.

Objetivos

Esta pesquisa tem como objetivo principal identificar os FA que possuem maior influência sobre o número de casos de dengue (NCD) registrados no município de Viçosa, tendo apenas o custo computacional para o processamento de dados públicos e economizando recursos com pesquisas de campo extensas.

Conclusões

Conclui-se que os dados públicos podem ser suficientes para a modelagem do problema e que embora o período de estudos seja curto, os resultados foram satisfatórios e o KNN apresentou um modelo que se adequa ao NCD, além disso indica-se a necessidade de relacionar mais FA e a posição geográfica para cada caso registrado, expandir a série temporal para análises mais consistentes e acuradas e se possível realizar um estudo de campo para coleta de ovos e larvas do mosquito, a fim de estabelecer uma relação entre os FA estudados, os ovos e os casos registrados com a devida defasagem de tempo entre as ocorrências para prever momentos de crises epidêmicas.

Material e Métodos

Os dados base para a pesquisa foram os NCD diários para o ano 2013 e FA descritos a seguir: precipitação total (PPTG4), temperatura média (TMG4), temperatura de orvalho média (TOMG4), temperatura mínima (TMNG4) e temperatura de orvalho mínima (TONG4). Todos os dados foram padronizados e os FA foram defasados em 1, 2, 3, 4, 5, 6, 7 e 8 semanas para que o não-sincronismo entre os FA e NCD fosse identificado, sendo que os melhores resultados de correlação se deram com a defasagem de 4 semanas. Utilizou-se as linguagens R e Python para o processamento dos dados.

Agradecimentos

Agradeço à comissão orientadora pela difusão do conhecimento ao longo da pesquisa, possibilitando a execução desta. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), mesmo sem vínculo direto com a pesquisa, fornece bolsas para manutenção dos pós-graduandos. Agradeço também à Universidade Federal de Viçosa, que em seus 96 anos tem sido uma instituição séria e comprometida com a ciência.