

TRIAGEM DE CASOS DE DENGUE UTILIZANDO DADOS CLÍNICOS E APRENDIZADO DE MÁQUINA

BIANCA CONRAD BOHM¹; SUELLEN CAROLINE MATOS SILVA²;
DANTON DIEGO FERREIRA³; FÁBIO RAPHAEL PASCOTI BRUHN⁴

¹*Universidade Federal de Pelotas – biankabohm@hotmail.com*

²*Universidade Federal de Pelotas – suellen.carol.as@gmail.com*

³*Universidade Federal de Larvas - danton@ufla.br*

⁴*Universidade Federal de Pelotas – fabio_rpb@yahoo.com.br*

1. INTRODUÇÃO

A dengue é uma doença reemergente que causa grande impacto na saúde pública, principalmente em países de clima tropical ou subtropical. É uma arbovírose transmitida pelos mosquitos do gênero *Aedes*, causada pelo vírus da família flaviviríde, classificado em 4 sorovares (DENV1, DENV2, DENV3 e DENV4) (SALIM et al., 2021; HUANG et al., 2020).

É caracterizada como uma doença febril aguda, dinâmica e sistêmica, sendo descrita com um ou mais dos seguintes sintomas: febre alta > 40°C, cefaleia, dor retro-orbital, náusea/vômito, mialgia, artralgia e erupções cutâneas (DAUMAS et al., 2013).

Pesquisadores elaboraram estratégias de diagnóstico baseado em resultados laboratoriais. Embora esses resultados sejam de extrema importância, em anos epidêmicos é inviável realizar o diagnóstico de todos os casos suspeitos, então muitos casos são confirmados apenas por critérios clínico-epidemiológicos. (TANNER et al., 2008; DAUMAS et al., 2013; DAVI et al., 2019; HUANG et al., 2020). Diante desse contexto, dispor de um sistema que detecta casos de dengue de forma precoce faz-se necessário em países com número alto de casos por ano, como o Brasil (DAVI et al., 2019). Além disso, o sistema de vigilância carece de ferramentas de fácil aplicação para realizar a triagem de pacientes de forma eficiente e assim, otimizar os atendimentos médicos e os recursos na saúde. Uma das técnicas que vem sendo explorada em diversas áreas de pesquisa, com resultados satisfatórios, é o aprendizado de máquina, uma metodologia que demonstra ser uma ferramenta eficaz para os serviços de saúde. Há diversas metodologias aplicáveis, de acordo com o problema em questão e a resposta esperada (DAVI et al., 2019; NEJAD & VARATHAN, 2021).

O aprendizado de máquina é uma técnica composta por algoritmos capazes de “aprender” com base nos seus erros e identificar quais os atributos do banco são mais relevantes. Os modelos de aprendizado de máquina utilizam algoritmos de aprendizagem que após um treinamento prévio do modelo tornam-se capazes de classificar os casos, agrupar dados em conjuntos semelhantes ou fazer a predição de valores (FAVAN et al., 2020).

Este trabalho tem como objetivos: 1) identificar características e sinais clínicos relevantes para realizar a triagem de casos de dengue; 2) avaliar a precisão dos modelos com e sem a combinação dessas características.

2. METODOLOGIA

Trata-se de uma pesquisa quantitativa, realizada com dados secundários provenientes das fichas de notificação individual do Sistema Nacional de Agravos

de Notificação (SINAN), obtidas no site DataSUS. Foi realizado o download das fichas de notificação de dengue dos estados de Minas Gerais e do Rio de Janeiro, nos anos de 2016 a 2019.

As variáveis pré-selecionadas foram: sexo, raça, idade, sinais clínicos e forma de diagnóstico e a variável dependente foi a classificação final (caso de dengue confirmado ou descartado por exame laboratorial). O teste de qui-quadrado foi realizado para avaliar a associação entre as variáveis independentes e a variável dependente, sendo que somente aquelas associadas ($p<0,05$) com a dependente foram selecionadas para construção do modelo.

Após a seleção das variáveis de interesse foi realizada a limpeza do banco, sendo retirados todos os casos em que o preenchimento da ficha não foi completo. Apenas casos confirmados ou descartados por critério laboratorial foram utilizados. Neste estudo foi utilizado o modelo árvore de decisão, o qual é indicado para resolver problemas de classificação. O conjunto de dados foi dividido em duas partes: treinamento (70%) e teste (30%). Para realizar o treinamento foi utilizada a técnica de validação cruzada ou k-fold, sendo o tamanho 10 o tamanho do k. A eficácia do modelo no diagnóstico da dengue foi avaliada pelas métricas de acurácia, sensibilidade e especificidade. Todas as variáveis independentes testadas se mostraram associadas a incidência da dengue ($p<0,05$), portanto, foram incluídas no modelo.

Foram realizados dois testes com o modelo. Um modelo foi testado com variáveis sociais (raça, idade e sexo) e sinais clínicos e o outro modelo foi testado só com dados clínicos.

A linguagem utilizada para executar os experimentos de aprendizado de máquina foi o Python em conjunto com as bibliotecas Scikit-Learn, Pandas e Numpy.

3. RESULTADOS E DISCUSSÃO

Minas Gerais e Rio de Janeiro notificaram 882.612 casos suspeitos de dengue. Destes foram 645.264 (73,11%) casos foram confirmados através de critérios clínico epidemiológico, 229.113 (21,21%) por critério laboratorial e em 8.235 casos o campo estava em branco ou com falha no preenchimento. Neste mesmo período 321.411 casos foram descartados, sendo que 135.163 casos foram descartados por critério laboratorial. A primeira limpeza do banco excluiu todos os casos confirmados por critérios clínicos epidemiológicos e os em branco. Após foram excluídos todos os casos em que continham dados ausentes nas variáveis de interesse.

As duas técnicas testadas (modelo com dados clínicos e sociais X modelo com dados clínicos) apresentaram resultados satisfatórios, com valores de acurácia acima de 90%. Porém o modelo de árvore aleatória treinado com dados dos indivíduos e sinais clínicos apresentou os melhores valores de acurácia, sensibilidade e especificidade, como apresentado na tabela 1.

Tabela 1: Resultados das métricas de avaliação dos modelos de aprendizagem de máquina utilizados para a triagem de casos de dengue.

Modelo	Variáveis	Acurácia	Sensibilidade	Especificidade
Árvore de decisão	Com dados sociais e clínicos	92,15	95,95	96,41

Só dados clínicos	90,86	87,00	95,68
-------------------	-------	-------	-------

A avaliação do modelo levou em conta os melhores resultados em termos de sensibilidade e especificidade e não apenas os valores de acurácia. A decisão de avaliar essas métricas deve-se ao fato de que a proposta do estudo é construir um modelo que auxilie na triagem de casos suspeitos, portanto, é importante que o modelo apresenta valores de sensibilidade elevados, a fim de minimizar a chance de se descartar um possível caso de dengue.

Daumas et al., (2009) encontraram resultados satisfatórios ao treinarem um algoritmo com dados clínicos e hematológicos. Apesar de mostrar bons resultados, utilizar dados laboratoriais acaba sendo um fator limitante para a implantação do modelo no Brasil, em função da elevada incidência observada em anos epidêmicos, tornando inviável a realização de exames laboratoriais para todos os casos suspeitos. Devido ao aumento na procura de atendimento médico e na incapacidade de realizar a confirmação laboratorial, o Ministério da saúde orienta que, em anos epidêmicos, o caso pode ser confirmado através de critérios clínico-epidemiológicos (MINISTÉRIO DA SAÚDE, 2013).

A cada ano, mais regiões notificam casos de dengue, revelando que o vetor está adaptado ao clima tropical e em expansão pelo território brasileiro (CUNHA et al., 2021). Visto isso, estratégias de triagem baseadas em evidências que identificam indivíduos com probabilidade de ter infecção por dengue nos estágios iniciais da doença são necessárias para auxiliar profissionais da saúde em investigações clínicas, além de contribuir para o gerenciamento e planejamento de recursos de saúde (SILITONGA et al., 2021).

Os serviços de saúde geram dados sobre atendimento médico, mas carecem de ferramentas que aprimorem o diagnóstico e a vigilância de doenças em áreas endêmicas. (DAUMAS et al., 2009). Porém, para esse tipo de trabalho é preciso lidar com as restrições de se trabalhar com dados secundários, como a presença de dados incompletos, o que consequentemente resulta em uma perda excessiva de dados, tornando-se um fator limitante para a obtenção de modelos mais precisos (CAMPOS et al., 2019). Uma alternativa seria a implantação de um recurso tecnológico como um aplicativo para *smartphone* onde se seleciona os sinais que o paciente apresenta e o aplicativo retorne com a classificação – dengue ou descartado.

4. CONCLUSÕES

O modelo de classificação foi eficaz na triagem de casos de dengue utilizando dados clínicos, assim, modelos de aprendizagem surgem como uma ferramenta com potencial para auxiliar na tomada de decisão médica.

5. AGRADECIMENTOS

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

6. REFERÊNCIAS BIBLIOGRÁFICAS

Salim, N., Wah, Y. B., Reeves, C., Smith, M., Yaacob, W., Mudin, R. N., Dapari, R., Sapri, N., & Haque, U. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. *Scientific reports*, v.11, n.1, p.939, 2021. <https://doi.org/10.1038/s41598-020-79193-2>

Huang, S. W., Tsai, H. P., Hung, S. J., Ko, W. C., & Wang, J. R. (2020). Assessing the risk of dengue severity using demographic information and laboratory test results with machine learning. *PLoS neglected tropical diseases*, v. 14, n.12, 2020.

Daumas, R.P., Passos, S.R., Oliveira, R.V. et al. Clinical and laboratory features that discriminate dengue from other febrile illnesses: a diagnostic accuracy study in Rio de Janeiro, Brazil. *BMC Infect Dis* v.13, n.77, 2013.

Tanner, L., Schreiber, M., Low, J. G., Ong, A., Tolfvenstam, T., Lai, Y. L., Ng, L. C., Leo, Y. S., Thi Puong, L., Vasudevan, S. G., Simmons, C. P., Hibberd, M. L., & Ooi, E. E.. Decision tree algorithms predict the diagnosis and outcome of dengue fever in the early phase of illness. *PLoS neglected tropical diseases*, v.2, n.3, 2008.

Davi, C. et al., "Severe Dengue Prognosis Using Human Genome Data and Machine Learning," in **IEEE Transactions on Biomedical Engineering**, v. 66, n. 10, pp. 2861-2868, 2019.

Nejad, F. Y., Varathan, K.D. Identification of significant climatic risk factors and machine learning models in dengue outbreak prediction. *BMC Med Inform Decis Mak* v.21, n.141, 2021. <https://doi.org/10.1186/s12911-021-01493-y>

Favan, J. R., dos Santos Coscolin, R. B., Jim, A. S., Gomes, R. L., & de Souza Passos, J. R. Modelos computacionais e estatísticos para a predição da severidade da mancha foliar causada por Xanthomonas spp. em clone híbrido de Eucalyptus grandis x Eucalyptus urophylla. *Tekhne e Logos*, v. 11, n.2, p.50-65, 2020

SINAN. **Sistema de Informação de Agravos de Notificação. 2015.** Disponível em: <http://portalsinan.saude.gov.br/images/documentos/Agravos/Dengue/DIC_DADOS_ONLINE.pdf> Acesso em: 17 abr. 2019.

Silitonga, P., Bustamam, A., Muradi, H., Mangunwardoyo, W., & Dewi, B. E. Comparison of Dengue Predictive Models Developed Using Artificial Neural Network and Discriminant Analysis with Small Dataset. *Applied Sciences*, v.11, n.3, p.943, 2021

Brasil. **Ministério da Saúde.** Secretaria de Vigilância em Saúde. Departamento de Vigilância das Doenças Transmissíveis. Dengue : diagnóstico e manejo clínico : adulto e criança [recurso eletrônico] / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância das Doenças Transmissíveis. – 5. ed. – Brasília : Ministério da Saúde, 2016. 58 p. : il.

Cunha, M. D. C. M., Ju, Y., Morais, M. H. F., Dronova, I., Ribeiro, S. P., Bruhn, F. R. P., Caiaffa, W. T. Disentangling associations between vegetation greenness and dengue in a Latin American city: Findings and challenges. *Landscape and urban planning*, v.216, 2021.

Campos NBD, Morais MHF, Ceolin APR, Cunha MdaCM, Nicolino RR, Schultes OL, FricheAAdeL, Caiaffa WT. Twenty-Two years of dengue fever (1996-2017): naepidemiologicalstudy in a Brazilian city, *International Journal of Environmental Health* 2019