

VALERIA: Uma Plataforma para Auxiliar o Diagnóstico e o Monitoramento de Arboviroses

Virtual Assistant for LEarning pRocesses In Arbovirus

Thomás Tabosa de Oliveira
Sebastião Rogério da Silva Neto
Igor Vitor Teixeira
Patricia Takako Endo
tto@ecomp.poli.br
srsn@ecomp.poli.br
ivt@ecomp.poli.br
patricia.endo@upe.br
Universidade de Pernambuco
Recife, Pernambuco, Brasil

Vanderson Souza Sampaio
vandersons@gmail.com
Fundação de Medicina Tropical Dr. Heitor Vieira Dourado
Fundação de Vigilância em Saúde Dra. Rosemary Costa
Pinto
Manaus, Amazonas, Brasil

ABSTRACT

Arboviruses are Neglected Tropical Diseases (NTDs) that mainly affect developing countries and Brazil is one of them. Such diseases have a negative impact on quality of life of the patients. One of the main challenges in combating these diseases is the difficulty of an accurate diagnosis among the concomitant arborival diseases due the similarity of symptoms. This project proposes VALERIA, a low-cost platform based on machine learning (ML) models for assisting the diagnose of arboviruses in the national scenario. VALERIA is proposed to improve the quality of life of patients, allowing the most appropriate treatment after a diagnose and therefore presenting a great potential of generating local public policies of social impact. A prototype was already developed for diagnosis and is in validation phase by public health specialists.

KEYWORDS

valeira, virtual assistant, arbovirus, machine learning, diagnose

1 INTRODUÇÃO

A Agenda 2030 das Organização das Nações Unidas (ONU)¹ é um compromisso feito pelos 193 Estados-membros para “*erradicar a pobreza e promover vida digna para todos, dentro dos limites do planeta*” com base em 17 Objetivos de Desenvolvimento Sustentável (ODS). Especificamente o ODS 3 - Saúde e Bem-estar possui a meta 3.3, “*até 2030, acabar com as epidemias de AIDS, tuberculose, malária e Doenças Tropicais Negligenciadas (DTNs), e combater a hepatite, doenças transmitidas pela água, e outras doenças transmissíveis*”.

As DTNs são um conjunto de doenças que tem uma prevalência maior em países tropicais e em desenvolvimento e estão associados a fatores como pobreza e baixo Índice de Desenvolvimento Humano (IDH) nas regiões [5]. O Brasil é um dos países com maior

quantidade de casos de DTNs, sendo as regiões Norte e Noroeste as mais afetadas [6]. As arboviroses, que fazem parte das DTNs, mais comuns no país são Dengue, Chikungunya e Zika, e todas elas são transmitidas pelos mosquitos *Aedes Aegypti* e *Aedes Albopictus*. O Brasil é, inclusive, o que apresenta o maior número de casos absolutos de Dengue nos últimos anos nas Américas [7].

A infecção por arbovírus pode apresentar alguns sintomas que podem persistir durante meses e até mesmo anos (como a artralgia no caso da Chikungunya), diminuindo a qualidade de vida dos pacientes, com impactos econômicos significativos, devido à redução da produtividade [3]. Especificamente, é válido ressaltar o grande impacto negativo na qualidade de vida da família e da criança diagnosticada com Síndrome Congênita do Zika Vírus (SCZV).

Assim, a circulação concomitante de diferentes arbovírus em território brasileiro demanda novos procedimentos destinados ao diagnóstico diferencial das arboviroses no contexto da vigilância epidemiológica. Porém, as limitações de ordem financeira derivadas dos elevados custos para realização dos exames laboratoriais de confirmação em larga escala, apresentam-se como um grande desafio para a garantia da acurácia das notificações dos casos da doença.

O método mais comumente utilizado de diagnóstico é com base nos quadro epidemiológico local. Testes rápidos existem, porém são eficazes apenas até o terceiro dia após o surgimento dos sintomas; após este prazo é necessário testes laboratoriais e sorológicos. Porém, alguns testes sorológicos podem apresentar reação cruzada com outras doenças [4], enquanto que os testes laboratoriais necessitam de equipamentos específicos, que além de demandarem custo (de implementação e operacional), podem não estar disponíveis em diversas cidades do Brasil.

Modelos de *Machine Learning* (ML) se apresentam como uma solução acessível e de baixo custo para auxiliar profissionais de saúde a realizarem um diagnóstico diferencial de arbovirose mais acurado no momento do atendimento do paciente. Os modelos de ML são capazes de reconhecer padrões a partir de uma base de dados e, assim, seria possível utilizar tais modelos para auxiliar no diagnóstico de um paciente utilizando apenas dados clínicos e sociodemográficos (sem a necessidade de exames complementares). Além disso, esses modelos, após a fase de treinamento, podem realizar a

¹<http://www.agenda2030.com.br>

classificação requerendo baixo poder computacional, podendo ser embutido em *tablets*, celulares ou sistemas *web*, por exemplo.

Este projeto é transdisciplinar e conta com pesquisadores da área de computação e saúde pública para desenvolver a VALERIA, uma plataforma de baixo custo para auxílio no diagnóstico diferencial de Dengue, Chikungunya e Zika utilizando apenas dados clínicos e sociodemográficos. A VALERIA, além de melhorar a qualidade de vida dos pacientes que tiverem seus tratamentos corretamente direcionados de acordo com seu diagnóstico, apresenta um grande potencial de gerar políticas públicas de impacto social.

2 VALERIA

A VALERIA (acrônimo para Assistente Virtual para Processos de Aprendizagem com Arboviroses, do inglês *Virtual Assistant for LEarning pRocesses In Arbovirus*) é uma plataforma para auxílio no diagnóstico clínico e monitoramento de arbovírus utilizando modelos de ML, para uso social, especialmente nas unidades de saúde do SUS. A Figura 1 apresenta as funcionalidade da plataforma da VALERIA, onde dois módulos básicos são propostos: módulo de diagnóstico e módulo de monitoramento.

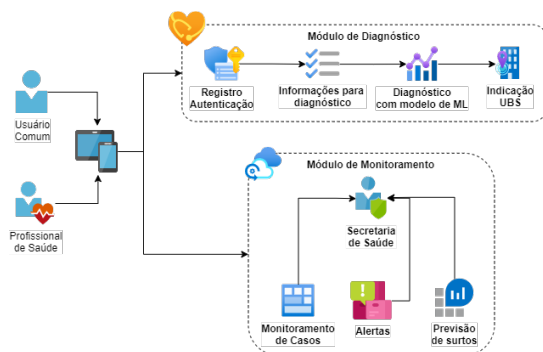


Figure 1: Funcionalidades da plataforma de diagnóstico clínico e monitoramento de arbovírus - VALERIA.

O módulo de diagnóstico diferencial de arbovirose é o principal módulo da VALERIA. Para utilizar o módulo, o profissional de saúde deve primeiramente se cadastrar e realizar sua autenticação no sistema. Ao acessar a VALERIA, o usuário poderá preencher um formulário simples, com as informações necessárias para realização do diagnóstico, utilizando um conjunto de onze sintomas (febre, mialgia, cefaleia, enxantema, náusea, dor nas costas, conjuntivite, artrite, artralgia, petéquias, dor retroorbital), e duas comorbidades prévias (diabetes e hipertensão). A partir dessas informações, o modelo de ML realiza o diagnóstico, informando as probabilidades do paciente possuir alguma das arboviroses: Dengue, Chikungunya, ou Zika, ou se o resultado é inconclusivo. Para cada uma dessas possíveis saídas é apresentada a taxa de confiança do modelo. É válido ressaltar que os sintomas e as comorbidades foram selecionadas através de uma extensa bateria de experimentos, seguindo metodologias rígidas e com auxílio de colaboradores especialistas em arboviroses.

O módulo do monitoramento apresenta um *dashboard* para visualização dos casos de arboviroses confirmados por município do país. Esse módulo conta com a integração com a Secretaria de Saúde

para geração de alertas endêmicos, bem como monitoramento em tempo real dos casos de uma determinada região. A geração de alertas é feita por modelos de ML com o objetivo de auxiliar as Secretarias de Saúde a definir um planejamento a médio e longo prazo de combate às arboviroses.

Como base de dados, esse módulo utilizará a base de dados do Sistema de Informação de Agravos e Notificação (SINAN)², uma base de dados nacional que contém dados de notificações de pacientes que foram diagnosticados com alguma doença da lista nacional de doenças de notificação compulsória³.

3 RELEVÂNCIA, IMPACTO SOCIAL E EXECUTABILIDADE DO PROJETO

Segundo os dados obtidos da Organização Pan Americana de Saúde, do inglês *Pan American Health Organization (PAHO)*⁴, em 2020, foram registrados 1.467.142 de casos prováveis de Dengue, 98.177 de Chikungunya e 18.941 de Zika no Brasil. Ainda segundo os relatórios do PAHO, foi possível observar que apenas 765.144 (52,15%) das notificações de Dengue, 39.461 (40,19%) das notificações de Chikungunya e 2.644 (13,95%) das notificações de Zika do Brasil foram confirmadas.

Considerando-se as limitações de natureza operacional relacionadas à confirmação diagnóstica em todo o Brasil, especialmente em períodos epidêmicos, assim como em áreas de acesso restrito, o desenvolvimento da VALERIA, composta por modelos computacionais para classificação diagnóstica baseada em sintomas clínicos, apresenta-se como uma solução de baixo custo que pode contribuir para o aprimoramento da acurácia das notificações de casos de Dengue, Chikungunya, Zika e também outras arboviroses. Essas doenças causam diversos impactos negativos na qualidade de vida do paciente infectado (como redução de produtividade, sintomas debilitantes e persistentes por longo período de tempo), tendo indicador de anos potenciais de vida perdidos (DALY) significativo, podendo chegar em casos fatais [3].

Este projeto de pesquisa está diretamente relacionado a Estratégia de Desenvolvimento e Carteiras de Projetos Públicos em Pernambuco para o período 2015/2035 (PE 2035)⁵, uma vez que o objetivo é oferecer tratamento preliminar de endemias e diminuição dos custos com saúde, e possui potencial para gerar políticas públicas locais de impacto social.

Apesar de apresentar inicialmente aplicabilidade no contexto dos Estados de Pernambuco e do Amazonas, a relevância e o impacto social deste projeto é nacional, sendo de fácil replicabilidade uma vez que a base de dados utilizada é proveniente do SINAN/SUS, e de baixo custo de implementação. O objetivo é prover melhor qualidade de vida para a população mais pobre, que geralmente é atendida em postos de saúde, onde muitas vezes não há recursos suficientes para o diagnóstico apropriado.

Além disso, considerando os parceiros envolvidos (ver sub-seção 4.1), a executabilidade deste projeto (implementação, teste, avaliação e validação da VALERIA) está dentro do plano inicial, com

²<http://portalsinan.saude.gov.br/>

³https://bvsms.saude.gov.br/bvs/saudelegis/gm/2020/prt0264_19_02_2020.html

⁴<https://www3.paho.org/data/index.php/en/mnu-topics.html>

⁵<https://www.comissaodaverdade.pe.gov.br/index.php/estrategia-de-desenvolvimento-e-carteira-de-projetos-publicos-pernambuco-2035>

recursos advindos de diferentes fontes. O projeto conta com a participação de pesquisadores de graduação e pós graduação que interagem diretamente com profissionais de saúde do sistema local. Esses profissionais trabalham diariamente no contexto do projeto (e em região de difícil acesso a recursos, como Internet) e assim, a problemática posta aqui é bastante realista e a solução é validada por especialistas na área.

A tecnologia aplicada, tanto do ponto de vista laboratorial, quanto de análises poderá ser transferida para a atenção primária e hospitalar, podendo ser aplicada na rotina das atividades da vigilância em saúde, contribuindo para a consolidação do SUS.

4 RESULTADOS OBTIDOS E ESPERADOS

O projeto ainda se encontra em fase de desenvolvimento. A proposta inicial da plataforma foi apresentada em [1] e o desenvolvimento e a avaliação do modelo ML utilizado para a classificação está em fase de revisão [2]. Atualmente um protótipo já foi desenvolvido e está em fase de validação por parte dos colaboradores especialistas na área de saúde pública. O protótipo foi desenvolvido na linguagem Python, com a utilização da biblioteca *streamlit*⁶ e pode ser acessado pelo *link*: <http://valeria.upecaruaru.com.br>. Um vídeo explicativo também está disponível no YouTube: <https://www.youtube.com/watch?v=3ipbYO4zJHo>.

Devido a ausência de dados de notificações de Zika nas bases de dados que foram utilizadas no treinamento do modelo de ML [2], o protótipo se limita a Dengue, Chikungunya e casos Inconclusivo. Foram avaliados sete modelos de ML para este protótipo (*Adaptive Boosting* (Adaboost), *Random Forest* (RF), *Gradient Boosting Machines* (GBM), *eXtreme Gradient Boosting* (Xgboost), *k-Nearest Neighbours* (KNN), *Naive Bayes* (NB) e *Multilayer Perceptron* (MLP)). Para o treinamento, foram utilizados dados de notificação de Dengue e Chikungunya do Estado do Amazonas e da Cidade do Recife, Pernambuco, de 2015 a 2020. Após o pré-processamento da base de dados, a mesma resultou em 17.172 registros e 27 atributos, divididos igualmente para cada uma das três classes. Técnicas de *feature selection* e *grid search* foram aplicadas para selecionar atributos relevantes e as melhores configurações dos modelos de ML. Atualmente, o modelo do protótipo é o *Gradient Boosting* com a seguinte

configuração: *max_depth* : 3 e *n_estimators* : 200. Após validação com especialista, 14 atributos foram selecionados, um informacional sobre o tempo que o paciente apresenta sintomatologia; 11 atributos referentes a sintomas do paciente; e 2 atributos informando a existência de comorbidade prévia.

A Figura 2a apresenta a tela principal para a realização de um diagnóstico. Caso o usuário possua alguma dúvida sobre o significado de algum termo, é possível visualizar a sua definição, como apresentado na Figura 2b. Após a realização do diagnóstico, a VALERIA apresenta o resultado, bem como uma tabela com as probabilidades de cada saída do modelo, para uma análise mais detalhada do diagnóstico (Figura 2c). No módulo de monitoramento (Figura 2d) é possível visualizar um simples *dashboard* com a evolução dos casos de Dengue, Chikungunya e casos Inconclusivos durante um determinado período de anos em todo território brasileiro.

O protótipo foi desenvolvido utilizando as diretrizes das 10 heurísticas de Nielsen⁷ para o desenvolvimento de uma interface amigável para os usuários, a saber: *Visibilidade do status do sistema*: Através do menu na barra lateral da plataforma, o usuário fica ciente da sua localização dentro do sistema, sendo informado em qual menu o mesmo se encontra. *Compatibilidade entre o sistema e o mundo real*: O sistema possui uma linguagem comum para que o usuário não tenha dificuldade na navegação. No momento do diagnóstico, são utilizados termos técnicos dos sintomas, porém os mesmo possuem caixa de ajuda para uma definição simples de cada sintoma. *Controle e liberdade para o usuário*: No protótipo não existe nenhum tipo de registro de informações. Caso o usuário preencha informações erradas durante o diagnóstico, o mesmo pode simplesmente preencher as informações novamente. *Consistência e Padronização*: As telas presentes no protótipo são simples e com poucos textos, com indicações diretas do que deve ser feito, além de que todas as telas seguem o mesmo padrão de visualização. *Prevenção de erros*: Como o sistema não possui nenhum registro de informação, e o diagnóstico é apenas informativo, não existem problemas técnicos caso o usuário informe os dados erroneamente. *Reconhecimento em vez de memorização*: Todas as telas do protótipo são simples e autoexplicativas, facilitando a memorização dos usuários durante a utilização. As caixas de texto auxiliam em

⁶<https://streamlit.io>

⁷<https://www.nngroup.com/articles/ten-usability-heuristics/>

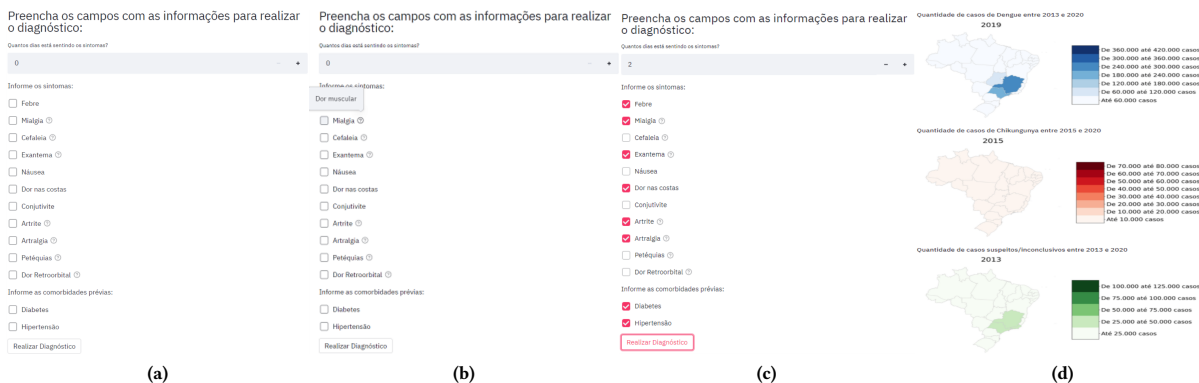


Figure 2: Principais telas da plataforma VALERIA.

quaisquer dúvidas que o usuário tenha a respeito das informações. *Eficiência e flexibilidade de uso*: Por possuir informações limitadas, o protótipo não possui nenhuma forma de customização ou de criação de atalhos para usuários, sendo um ponto de melhoria no futuro. *Estética e design minimalista*: O design minimalista do protótipo possui apenas as informações essenciais para o usuário. *Ajude os usuários a reconhecerem, diagnosticarem e recuperarem-se de erros*: Como as funcionalidades do protótipo são de caráter informativo, não se faz necessário a criação de mensagem de erros. *Ajuda e documentação*: A tela de ajuda do protótipo possui um vídeo de curta duração explicando como o diagnóstico pode ser realizado na plataforma para auxiliar o usuário.

Assim, uma vez que os resultados desse estudo pretendem melhorar a capacidade de classificação das diferentes arboviroses existentes, esta pesquisa pode contribuir para o manejo clínico mais adequado (melhoria da abordagem terapêutica) e o aprimoramento da vigilância de saúde, com geração de propostas de políticas públicas na área. Também é esperado que os resultados do presente projeto tenham impacto no monitoramento de sinais e sintomas das doenças, implicando a redução dos danos colaterais aos pacientes; além de notificação prévia de possíveis epidemias, para que as devidas ações de prevenção possam ser tomadas em tempo hábil.

4.1 Recursos envolvidos e parceiros

Este projeto de pesquisa está sendo desenvolvido por pesquisadores do Programa de Pós-Graduação em Engenharia da Computação (PPGEC) da Universidade de Pernambuco (UPE): Thomás Tabosa (mestrando do PPGEC)⁸; Sebastião da Silva Neto (doutorando do PPGEC)⁹; Igor Teixeira (mestrando do PPGEC)¹⁰; e profa. Dra. Patricia Endo (coordenadora do projeto e professora adjunta da UPE, com larga experiência na área de computação em nuvem e aplicação de ML na área de saúde)¹¹ em parceria com pesquisadores do Programa de Pós-Graduação em Medicina Tropical da Universidade do Estado do Amazonas (UEA), do Programa de Pós-Graduação em Ciências da Saúde da Universidade Federal do Amazonas (UFAM), da Fundação de Medicina Tropical Dr. Heitor Vieira Dourado (FMT-HVD) e da Fundação de Vigilância em Saúde Dra. Rosemary Costa Pinto (FVS-RCP/AM), representado pelo professor Dr. Vanderson Sampaio¹², que tem colaborado com projetos envolvendo a aplicação de métodos computacionais para identificação de preditores de óbitos e desfechos clínicos desfavoráveis em doenças como arboviroses e tuberculose.

Este projeto possui recursos advindo dos seguintes editais:

- Edital CNPq/AWS 032/2019 - Acesso às Plataformas de Computação em Nuvem da AWS (Cloud Credits for Research), no valor de USD \$ 11.226 (em créditos em serviço de computação em nuvem da AWS);
- Edital 006/2019 referente ao Programa de Apoio à Pesquisa - Universal Amazonas da FAPEAM, no valor de R\$ 62.212,00;
- Edital de Apoio a Pós-Graduação Stricto Sensu da UPE 2020 - Modalidade Auxílio para Projetos de Pesquisa (APQ), no valor de R\$ 5.000,00;

⁸<http://lattes.cnpq.br/0487004776163889>

⁹<http://lattes.cnpq.br/5589837708731892>

¹⁰<http://lattes.cnpq.br/7616014062895959>

¹¹<http://lattes.cnpq.br/5055727404635243>

¹²<http://lattes.cnpq.br/0039836167659650>

- Edital 20/2019 referente a Concessão de Bolsas de Pós-Graduação Stricto Sensu da FACEPE - Bolsa de Mestrado e de Doutorado.

4.2 Cronograma

O seguinte cronograma define as atividades principais para os próximos passos no desenvolvimento e validação da VALERIA.

- 2021: Validação do protótipo por parte dos especialistas da área de saúde pública;
- 2022: Criação da versão *offline* da VALERIA;
- 2023: Proposição dos modelos de ML para classificação de Dengue, Chikungunya, Zika e Inconclusivo; e dos modelos para previsão temporal de ocorrência de casos de arboviroses;
- 2024: Implementação da plataforma do módulo de monitoramento; e teste e validação da VALERIA com todas as funcionalidades.

5 CONSIDERAÇÕES FINAIS

Este trabalho propõe a criação da VALERIA, uma plataforma para auxiliar no diagnóstico de baixo custo e no monitoramento de arboviroses no território brasileiro. A plataforma possui dois módulos principais: (a) diagnóstico, para a realização de um diagnóstico rápido para os pacientes no momento de seu atendimento no posto de saúde; e (b) monitoramento, para análise do quadro epidêmico local e auxiliar na tomada de decisão de campanhas e políticas públicas de combate às arboviroses.

6 AGRADECIMENTOS

Agradecemos ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a Fundação de Amparo à Pesquisa do Estado do Amazonas (FAPEAM), a Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) e a Universidade de Pernambuco (UPE).

REFERENCES

- [1] Sebastião Rogério da Silva Neto, Thomás Tabosa de Oliveira, Vanderson de Souza Sampaio, Theo Lynn, and Patricia Takako Endo. 2020. Platform for monitoring and clinical diagnosis of arboviruses using computational models. In *2020 International Conference on Cyber Security and Protection of Digital Services (Cyber Security)*. IEEE, 1–3.
- [2] Thomás Tabosa de Oliveira, Sebastião Rogério da Silva Neto, Igor Vitor Teixeira, Samuel Benjamin, Vanderson de Souza Sampaio, and Patricia Takako Endo. 2021. A comparative study of machine learning techniques for multi-class classification of arboviral diseases. *Under review* (2021).
- [3] Nildimar Alves Honório, Daniel Cardoso Portela Câmara, Guilherme Amaral Calvet, and Patricia Brasil. 2015. Chikungunya: uma arbovirose em estabelecimento e expansão no Brasil. *Cadernos de saude publica* 31 (2015), 906–908.
- [4] Tamara Nunes Lima-Camara. 2016. Arboviroses emergentes e novos desafios para a saúde pública no Brasil. *Revista de Saúde Pública* 50 (2016), 36.
- [5] José Angelo L Lindoso and Ana Angélica BP Lindoso. 2009. Doenças tropicais negligenciadas no Brasil. *Revista do Instituto de Medicina Tropical de São Paulo* 51, 5 (2009), 247–253.
- [6] Francisco Rogerlândio Martins-Melo, Alberto Novaes Ramos Jr, Carlos Henrique Alencar, and Jorg Heukelbach. 2016. Mortality from neglected tropical diseases in Brazil, 2000–2011. *Bulletin of the World Health Organization* 94, 2 (2016), 103.
- [7] OPAS. 2020. Casos de dengue nas Américas chegam a 1,6 milhão, o que destaca a necessidade do controle de mosquitos durante a pandemia. <https://www.paho.org/pt/noticias/23-6-2020-casos-dengue-nas-americas-chegam-16-milhao-que-destaca-necessidade-do-controle>