

Predição de casos de arboviroses no Recife utilizando computação de reservatório

Arbovirus case prediction in Recife using reservoir computing

Predicción de casos de arbovirus en Recife utilizando computación de yacimientos

Ana Clara Gomes da Silva¹, Cláudia Priscila Nunes Silva², Clarisse Lins de Lima³, Danilo Wanderley Lapa⁴, Felipe Estevão da Silva⁵, Mariana Marinho da Silva Andrade⁴, Arianne Sarmiento Torcate⁶, Cecília Cordeiro da Silva⁷, Giselle Machado Magalhães Moreno⁸, Wellington Pinheiro dos Santos⁹

1 Mestra em Engenharia Biomédica, Universidade de Pernambuco, Recife (PE), Brasil.

2 Mestra em Matemática, Universidade de Pernambuco, Recife (PE), Brasil

3 Doutora em Engenharia da Computação, Universidade Federal de Pernambuco, Recife (PE), Brasil

4 Graduando em Sistemas de Informação, Universidade Federal de Pernambuco, Recife (PE), Brasil

5 Graduando em Engenharia Eletrônica, Universidade Federal de Pernambuco, Recife (PE), Brasil

6 Mestra em Engenharia da Computação, Universidade de Pernambuco, Recife (PE), Brasil

7 Doutora em Engenharia da Computação, Universidade Federal de Pernambuco, Recife (PE), Brasil

8 Doutora em Neurociências, Universidade Federal de Pernambuco, Recife (PE), Brasil

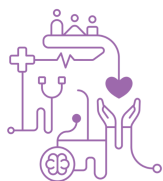
9 Doutor em Engenharia Elétrica, Universidade Federal de Pernambuco, Recife (PE), Brasil

Autor correspondente: Prof. Dr. Wellington Pinheiro dos Santos

E-mail: wellington.santos@ufpe.br

Resumo

Objetivo: Devido à complexidade do diagnóstico das arboviroses, a predição com aprendizado de máquina objetiva antecipar surtos, trazer rapidez para o tratamento e diminuir a propagação. **Método:** Este estudo propõe aplicar técnicas de computação de reservatório, incorporando fatores climáticos, para prever surtos e aumentos na incidência dessas doenças. **Resultados:** Os modelos criados tiveram valores maiores de 0,80 de acurácia, *precision* e *recall* na previsão de casos no Recife, Pernambuco. **Conclusão:** Esses modelos são cruciais para tomada de decisão, permitindo intervenções preventivas e de controle mais eficazes contra arboviroses na saúde pública.



Descritores: Predição de arboviroses; Modelo de predição; Computação de reservatório

Abstract

Objective: Due to the complexity of diagnosing arboviruses, machine learning prediction aims to anticipate outbreaks, speed up treatment and reduce spread. **Method:** This study proposes applying reservoir computing techniques, incorporating climatic factors, to predict outbreaks and increases in the incidence of these diseases. **Results:** The models created had values greater than 0.80 for accuracy, precision and recall in predicting cases in Recife, Pernambuco. **Conclusion:** These models are crucial for decision-making, enabling more effective preventive and control interventions against arboviruses in public health.

Keywords: Arbovirus Forecasting; Forecasting Model; Reservoir computing

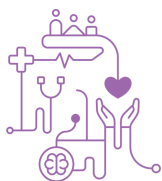
Resumen

Objetivo: Debido a la complejidad del diagnóstico de los arbovirus, la predicción mediante aprendizaje automático pretende anticipar los brotes, acelerar el tratamiento y reducir la propagación. **Método:** Este estudio propone aplicar técnicas de computación de reservorios, incorporando factores climáticos, para predecir brotes y aumentos en la incidencia de estas enfermedades. **Resultados:** Los modelos creados tuvieron valores superiores a 0,80 para exactitud, *precision* y *recall* en la predicción de casos en Recife, Pernambuco. **Conclusión:** Estos modelos son cruciales para la toma de decisiones, permitiendo intervenciones preventivas y de control más eficaces contra los arbovirus en la salud pública.

Descriptores: Predição de arbovirose; Modelo de predicción; Computación de yacimientos

Introdução

As arboviroses, representadas pela dengue, zika e chikungunya, são doenças transmitidas pelo mosquito *Aedes aegypti* e têm se tornado um desafio crescente para a saúde pública global. Estima-se que a dengue resulte em até 390 milhões de infecções anuais em todo o mundo^(1,2). A zika ganhou destaque global devido à epidemia na América Latina em 2015, que foi associada a casos de microcefalia em recém-nascidos⁽³⁾.

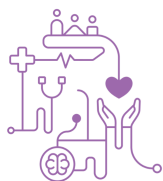


A chikungunya teve um aumento significativo de casos nas últimas décadas, com surtos em várias partes do mundo⁽⁴⁾.

Essas arboviroses trazem prejuízos significativos para o Brasil e outros países tropicais e subtropicais, tanto de natureza econômica quanto de saúde pública. E essas doenças podem custar caro ao sistema de saúde em testes, consultas e medicamentos. As doenças também podem impedir a criação ou cortar mais de 214.000 empregos⁽⁵⁾.

As manifestações neurológicas causadas por esses vírus têm alto potencial de morte ou sequelas. As complicações que ocorrem no sistema nervoso associadas às arboviroses podem representar um desafio diagnóstico e de tratamento⁽⁶⁾. Além disso, a coexistência de epidemias, especialmente em áreas urbanas, aumenta a demanda por serviços de saúde. O diagnóstico é desafiador mesmo quando os testes confirmatórios estão disponíveis, pois dengue, chikungunya, zika e COVID-19 podem apresentar manifestações clínicas e características laboratoriais semelhantes nos estágios iniciais das doenças^(7,8). Portanto, o diagnóstico incorreto pode atrasar o cuidado e o manejo apropriados da doença, resultando em maior disseminação da doença e piores resultados clínicos⁽⁹⁾.

Devido à presença do mesmo vetor para as três doenças, áreas afetadas por uma delas também estão em risco para as outras, destacando a importância de medidas preventivas abrangentes para o controle dessas arboviroses⁽⁷⁾. Diante desse contexto, surgem estratégias tecnológicas para antecipar surtos por meio do monitoramento, mapeamento e diagnósticos mais céleres, entre outras intervenções^(10,11). Essas abordagens tecnológicas são cada vez mais pertinentes e constantemente aprimoradas, buscando uma resposta mais eficaz às arboviroses⁽¹⁰⁾. Nesse sentido, propõe-se neste estudo a aplicação de técnicas de Machine Learning para prever surtos e aumentos na incidência dessas doenças, incorporando diversos fatores como dados climáticos, umidade, pluviometria e temperatura. A integração dessas informações, através de uma abordagem de geoprocessamento aliada a redes neurais de pesos aleatórios, almeja melhorar a compreensão e previsão de casos de arboviroses.

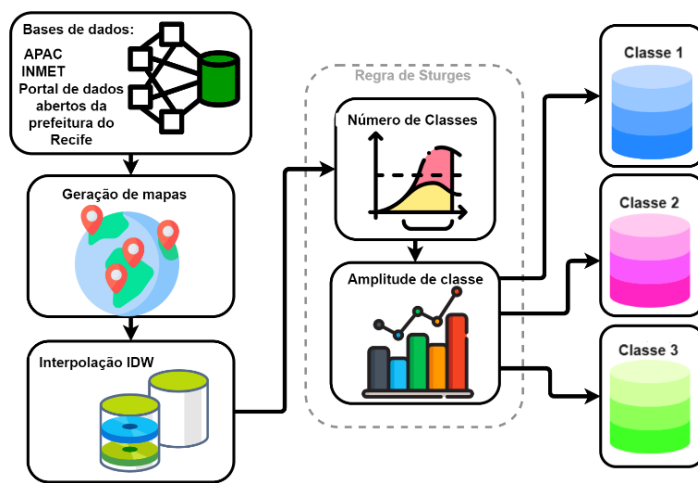


Métodos

Proposta de Aplicação

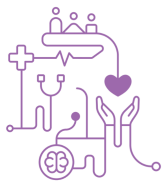
Na presente pesquisa, propomos uma análise espaço-temporal para a previsão da incidência de arboviroses, utilizando dados coletados de fontes públicas combinados com informações climáticas. Os dados foram adquiridos de plataformas de dados em âmbito nacional, estadual e municipal. Em seguida, os dados foram processados e convertidos para o formato .csv para uma melhor integração com as ferramentas de análise. Para a formação dos conjuntos de dados de treinamento, utilizamos um módulo de interpolação IDW(Inverse Distance Weighting) que usa o método da ponderação pelo inverso da distância. Dois algoritmos foram empregados com o objetivo de avaliar seu desempenho, mantendo as mesmas configurações de treinamento. Por fim, foi realizada uma comparação entre os modelos criados, buscando determinar qual apresenta o melhor desempenho nos critérios de avaliação escolhidos.

Figura 1 - Diagrama da metodologia usada para criação dos modelos e distribuídos em classes de acordo com a incidência de casos.



Base de Dados

Os dados empregados nos experimentos foram retirados de fontes públicas. As informações meteorológicas foram adquiridas de plataformas estaduais e nacionais, compreendendo dados mensais de temperatura máxima e velocidade do vento,

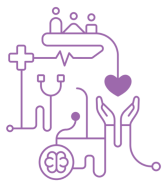


provenientes do Instituto Nacional de Meteorologia (INMET), referentes ao período entre 2014 e 2016.

Para os registros de precipitação, consultamos o Sistema Integrado de Gestão de Recursos Hídricos (SIGRH), desenvolvido pela Agência Pernambucana de Águas e Clima (Apac), que forneceu dados de chuvas diárias e acumuladas mensalmente em três postos pluviométricos na cidade de Recife.

Os registros de casos de arboviroses foram obtidos através do Portal de Dados Abertos da Prefeitura do Recife, gerenciado pela Empresa Municipal de Informática (EMPREL). Esses dados, que compreendem informações de saúde pública entre os anos de 2014 e 2016, preservam a identificação dos pacientes pela Secretaria de Saúde da Cidade do Recife. Foram utilizados 29 conjuntos de dados relacionados a unidades de saúde, contendo registros de dengue, zika e chikungunya, classificados por bimestre devido às ações de controle vetorial realizadas nessa periodicidade, conforme orientações do Ministério da Saúde. Cada conjunto conta com 15446 instâncias. Todos os conjuntos têm 44 atributos, onde a saída do conjunto é o valor do pixel no mapa de distribuição do criadouro na coordenada correspondente. Para a análise dos resultados, foram registradas informações de latitude, longitude, distribuição bimestral dos casos, temperatura, precipitação e velocidade do vento para cada mês correspondente bimestralmente. A previsão foi elaborada considerando informações dos seis bimestres anteriores. Feito isso, a quantidade ideal de classes foi determinada utilizando a Regra de Sturges, e em seguida, foi calculada a amplitude de cada classe obtida. Por fim, a técnica de balanceamento de classes SMOTE (Synthetic Minority Oversampling Technique) com $k=3$ foi aplicada no software Weka para equilibrar a distribuição das classes, criando novos exemplos sintéticos para a classe minoritária com base na interpolação entre exemplos existentes.

Para analisar os resultados, foram registradas as informações de latitude, longitude, distribuição bimestral dos casos, temperatura, precipitação e velocidade do vento para cada mês correspondente bimestralmente. A previsão foi elaborada considerando informações dos seis bimestres anteriores, e a Regra de Sturges⁽¹²⁾ foi



aplicada com o objetivo de encontrar a quantidade ideal de classes, a partir da seguinte fórmula:

$$ONC = 1 + 3,33 \log(N)$$

Posteriormente, foi calculada a amplitude de cada classe obtida, usando a equação⁽¹²⁾:

$$OCW = \frac{range}{ONC}$$

Em seguida, foi utilizada a técnica de balanceamento de classes SMOTE (Synthetic Minority Oversampling Technique) com $k=3$ no software Weka, criando novos exemplos sintéticos para a classe minoritária, com base na interpolação entre exemplos existentes, de modo a equilibrar a distribuição das classes.

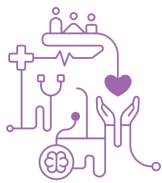
Modelos

Extreme Learning Machine (ELM)

Desenvolvido em 2006, o método de Extreme Learning Machine traz uma abordagem diferente de redes neurais, ao contrário do método iterativo de backpropagation usado em redes neurais convencionais, o ELM adota uma abordagem feedforward de treinamento único e simples ao atribuir pesos e vieses aleatórios entre a camada de entrada e a camada intermediária⁽¹³⁾. O ELM funciona em base da função de inversão de matriz de Moore-Penrose minimizando o erro quadrático médio entre o sinal que chega da camada intermediária e a camada de saída tratando a rede como um sistema linear⁽¹³⁾. As ELMs têm capacidade de generalização em comparação com outros métodos não-lineares.

Echo State Network (ESN)

Echo State Networks (ESN) é uma arquitetura de redes neurais de pesos aleatórios⁽¹⁴⁾. Os neurônios da camada oculta no tempo t recebem informações tanto do input no tempo t quanto da saída dos próprios neurônios do reservatório no tempo anterior ($t - 1$), conhecido como laço de realimentação⁽¹⁵⁾. Devido à natureza não linear das RNNs, elas são muito úteis para lidar com séries temporais epidemiológicas, como é o caso das arboviroses. A ESN é caracterizada por ter uma arquitetura composta por três camadas: (i) camada de inputs, (ii) camada de "reservatório" inicializada com um número aleatório de neurônios conectados, onde cada neurônio tem sua função de ativação, e (iii) camada de outputs⁽¹⁶⁾.



O modelo ESN demonstra habilidade para reter informações sobre recorrências temporais anteriores nos dados de entrada, graças ao seu reservatório. Ele também exibe não linearidade através das funções de ativação presentes nos neurônios da camada oculta, o que proporciona adaptabilidade e a capacidade de lidar com relações complexas.

Arquitetura dos modelos

Neste estudo, os modelos ESN e ELM foram desenvolvidos utilizando o *PyRCN* (Python Reservoir Computing Networks), um framework em Python que se baseia em bibliotecas dedicadas à computação científica, como *numpy* (Numerical Python) e *scipy* (Scientific Python), que permite otimizar, treinar e analisar *RCNs*⁽¹⁷⁾. Isso confere total compatibilidade do modelo com a biblioteca *scikit-learn*, permitindo a utilização direta dos métodos integrados de otimização de parâmetros desse framework. Foram criados modelos com 2, 5 e 10 camadas, contendo 10, 20, 50 e 100 neurônios em cada camada e cada configuração foi avaliada 30 vezes, utilizando validação cruzada com 10 *folds*.

Métricas

Neste trabalho, todos os modelos foram avaliados seguindo as métricas disponíveis da biblioteca *scikit-learn* do Python, utilizamos a acurácia, *precision* (precisão) e *recall* (sensibilidade), e são representadas pelas seguintes equações

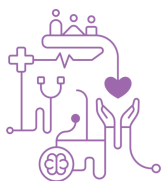
$$Acurácia = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = Sensibilidade = \frac{TP}{TP + FN} \quad (6)$$

Onde TP representa os verdadeiros positivos, TN os verdadeiros negativos, FP falsos positivos e FN falsos negativos⁽¹⁸⁾.

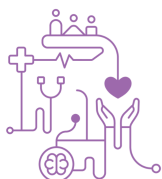
Resultados e Discussões



Nos modelos com arquitetura ELM, observamos que a configuração mais eficaz consistiu em duas camadas contendo 100 neurônios cada. Essa composição resultou em uma acurácia de 0,96231, precisão de 0,96234 e recall de 0,96285, com um tempo de treinamento notavelmente baixo de apenas 0,09 segundos. O desvio padrão das métricas permaneceu reduzido, variando entre 0,01924 e 0,10557. A segunda melhor configuração foi caracterizada por uma única camada composta por 100 neurônios, com uma acurácia de 0,96217, precisão de 0,96220, recall de 0,96294 e tempo de treinamento de 0,04 segundos. Contrariamente, a pior performance foi identificada na configuração com 10 camadas, cada uma contendo 10 neurônios. Os resultados incluíram uma acurácia de 0,62611, precisão de 0,62658 e recall de 0,63419, com um tempo de treinamento de 0,05 segundo.

Ao analisar as diferentes configurações de modelos, observou-se que aqueles com duas camadas escondidas e 100 neurônios apresentaram o melhor desempenho em termos de acurácia, *precision* e *recall*. No entanto, é importante notar que todos os modelos utilizando 100 neurônios, com 2, 5 e 10 camadas, alcançaram desempenhos semelhantes, com uma média acima 95% para acurácia e de 0,95 para *precision* e *recall* também. E mais, mais da metade dos modelos tiveram médias de acurácia acima de 92%, *precision* e *recall* acima de 0,92 também. Isso ressalta a importância de considerar não apenas a precisão, mas também o tempo de treinamento ao selecionar a configuração do modelo. Além disso, os tempos de treinamento são bem baixos, sendo 75% menores que 0,10 segundos.

Já nos modelos com arquitetura ESN, a configuração mais eficiente foi aquela com uma única camada composta por 100 neurônios. Nessa configuração, o modelo alcançou uma acurácia de 0,95438, precisão de 0,95422 e recall de 0,95513, com um tempo de treinamento de 0,30 segundos. Os desvios padrões das métricas permaneceram pequenos, variando entre 0,02082 e 0,02105. Em segundo lugar, a configuração de uma única camada com 20 neurônios apresentou métricas competitivas, com acurácia de 0,95249, *precision* de 0,95233 e *recall* de 0,95331, e um tempo de treinamento de 0,19 segundos. O desempenho menos satisfatório foi registrado no modelo com 10 camadas,

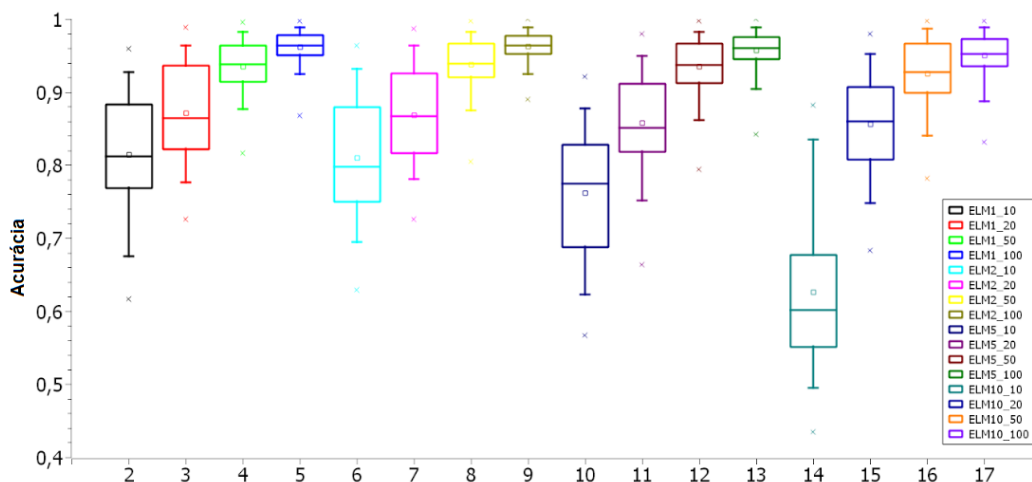


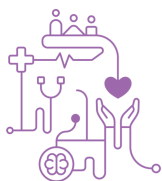
cada uma contendo 100 neurônios. Este apresentou uma acurácia de 0,61192, *precision* de 0,61146 e *recall* de 0,62677, com um tempo de treinamento substancialmente mais longo de 12,29 segundos.

Figura 1 – Tabela de resultados dos experimentos feitos com a arquitetura *Extreme Learning Machine*.

Configurações para a arquitetura ELM		Acurácia (%)	<i>Precision</i>	<i>Recall</i>	Tempo de treinamento (s)
Camada única	10 neurônios	81,41 ± 0,08	0,81 ± 0,08	0,81 ± 0,08	0,09 ± 0,02
	20 neurônios	87,14 ± 0,06	0,87 ± 0,06	0,87 ± 0,06	0,11 ± 0,00
	50 neurônios	93,56 ± 0,03	0,94 ± 0,03	0,94 ± 0,03	0,03 ± 0,01
	100 neurônios	96,22 ± 0,02	0,96 ± 0,02	0,96 ± 0,02	0,04 ± 0,01
2 camadas	10 neurônios	81,03 ± 0,08	0,81 ± 0,08	0,81 ± 0,08	0,02 ± 0,00
	20 neurônios	86,83 ± 0,06	0,87 ± 0,06	0,87 ± 0,06	0,03 ± 0,01
	50 neurônios	93,80 ± 0,03	0,94 ± 0,03	0,94 ± 0,03	0,05 ± 0,02
	100 neurônios	96,23 ± 0,02	0,96 ± 0,02	0,96 ± 0,02	0,09 ± 0,02
5 camadas	10 neurônios	76,20 ± 0,08	0,76 ± 0,08	0,76 ± 0,08	0,03 ± 0,01
	20 neurônios	85,79 ± 0,06	0,86 ± 0,06	0,86 ± 0,06	0,05 ± 0,01
	50 neurônios	93,51 ± 0,04	0,94 ± 0,04	0,94 ± 0,04	0,12 ± 0,03
	100 neurônios	95,73 ± 0,02	0,96 ± 0,02	0,96 ± 0,02	0,22 ± 0,05
10 camadas	10 neurônios	62,61 ± 0,10	0,63 ± 0,10	0,63 ± 0,11	0,05 ± 0,02
	20 neurônios	85,64 ± 0,06	0,86 ± 0,06	0,86 ± 0,06	0,08 ± 0,02
	50 neurônios	92,54 ± 0,04	0,93 ± 0,04	0,93 ± 0,04	0,23 ± 0,05
	100 neurônios	95,01 ± 0,03	0,95 ± 0,03	0,95 ± 0,03	0,47 ± 0,11

Figura 4 - Gráfico boxplot da acurácia dos modelos criados com o algoritmo *Extreme Learning Machine*.



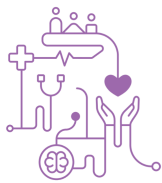


A utilização de gráficos boxplot da acurácia permitiu uma visualização clara da distribuição dos dados e do desempenho dos modelos. Pelo gráfico da Figura 4 é possível perceber que uma crescente nos valores de acurácia ao aumentar o número de neurônios nas camadas, além de diminuir a quantidade de dados que se diferenciam muito dos outros (outliers). Assim, o gráfico boxplot mostra que os modelos de uma, duas, cinco e dez camadas com 100 neurônios são, nesta ordem, os melhores resultados de acurácia.

Assim, ao analisar as diferentes configurações de modelos, observou-se que aqueles com duas camadas escondidas e 100 neurônios apresentaram o melhor desempenho em termos de acurácia, *precision* e *recall*. No entanto, é importante notar que todos os modelos utilizando 100 neurônios, com 2, 5 e 10 camadas, alcançaram desempenhos semelhantes, com uma média acima 95% para acurácia e de 0,95 para *precision* e *recall* também. E mais, mais da metade dos modelos tiveram médias de acurácia acima de 92%, *precision* e *recall* acima de 0,92 também. Isso ressalta a importância de considerar não apenas a precisão, mas também o tempo de treinamento ao selecionar a configuração do modelo. Além disso, os tempos de treinamento são bem baixos, sendo 75% menores que 0,10 segundos, o que é característico do algoritmo que foi usado no modelo.

De acordo com a análise das configurações usando o algoritmo ESN, as melhores configurações foram dos modelos com apenas uma camada tanto quanto as métricas usadas quanto ao tempo de treinamento. Contudo, dentre os modelos de camada única o com 100 neurônios apresentou o maior tempo de treinamento. Metade dos modelos criados com algoritmo ESN obtiveram valores de acurácia superior a 90%, *precision* e *recall* também foram superiores a 0,90. Além disso, apesar da grande variação dos tempos de treinamento, que foi de 0,09 a 12,29 segundos, mas metade estava abaixo de 0,72 segundos. O que faz sentido visto que foram justamente as configurações com 1 ou 2 camadas.

Figura 2 – Tabela dos resultados dos experimentos feitos com a arquitetura *Echo State Networks*.

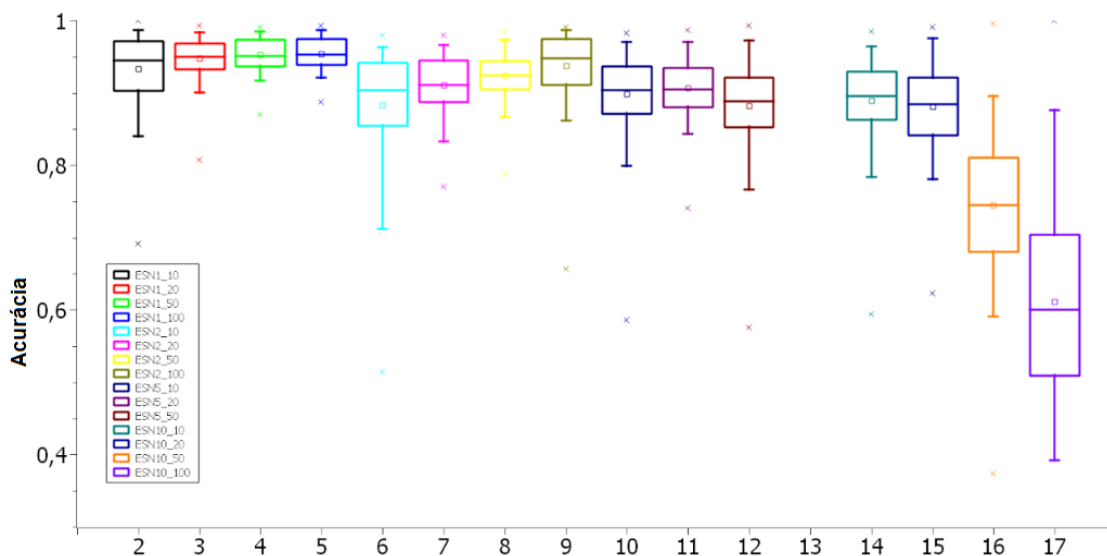
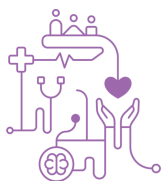


Configurações para a arquitetura ESN		Acurácia (%)	Precision	Recall	Tempo de treinamento (s)
Camada única	10 neurônios	93,29 ± 0,04	0,93 ± 0,05	0,93 ± 0,05	0,09 ± 0,02
	20 neurônios	94,81 ± 0,03	0,95 ± 0,03	0,95 ± 0,03	0,11 ± 0,03
	50 neurônios	95,24 ± 0,02	0,95 ± 0,02	0,95 ± 0,02	0,19 ± 0,05
	100 neurônios	95,44 ± 0,02	0,96 ± 0,02	0,96 ± 0,02	0,30 ± 0,07
2 camadas	10 neurônios	88,29 ± 0,08	0,89 ± 0,08	0,89 ± 0,08	0,43 ± 0,13
	20 neurônios	91,09 ± 0,04	0,91 ± 0,04	0,91 ± 0,04	0,49 ± 0,11
	50 neurônios	92,41 ± 0,03	0,93 ± 0,03	0,93 ± 0,03	0,85 ± 0,13
	100 neurônios	93,77 ± 0,04	0,94 ± 0,04	0,94 ± 0,04	0,71 ± 0,44
5 camadas	10 neurônios	89,90 ± 0,05	0,90 ± 0,05	0,90 ± 0,05	1,25 ± 0,52
	20 neurônios	90,68 ± 0,04	0,91 ± 0,04	0,91 ± 0,04	1,46 ± 0,30
	50 neurônios	88,22 ± 0,06	0,89 ± 0,05	0,90 ± 0,05	2,11 ± 0,34
	100 neurônios	75,86 ± 0,10	0,78 ± 0,09	0,78 ± 0,09	5,52 ± 134,41
10 camadas	10 neurônios	88,97 ± 0,06	0,89 ± 0,06	0,89 ± 0,06	2,46 ± 1,03
	20 neurônios	88,11 ± 0,06	0,89 ± 0,06	0,89 ± 0,06	2,23 ± 4,85
	50 neurônios	74,45 ± 0,09	0,77 ± 0,09	0,77 ± 0,09	2,78 ± 0,66
	100 neurônios	61,19 ± 0,14	0,63 ± 0,15	0,63 ± 0,15	12,29 ± 3,51

Pelo gráfico boxplot de acurácia dos modelos criados com algoritmo ESN, é possível ver que o modelo de camada única e 100 neurônios é o melhor, mais que isso, também é possível ver que os modelos de camada única têm maior precisão e baixa presença de dados incomuns (outliers) e diferem do gráfico dos outros modelos com mais de uma camada.

Assim, temos que de todos os modelos criados o melhor resultado foi criado com algoritmo ELM, de 2 camadas e 100 neurônios por camada, seguidos por modelos de camada única com 100 neurônios e duas camadas com 100 neurônios também criados com algoritmo ELM. E em quarto lugar temos o modelo criado com ESN de camada única com 100 neurônios. E todos esses com valores de tempo de treinamento muito baixos.

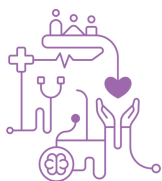
Figura 4 - Gráfico boxplot da acurácia dos modelos criados com o algoritmo *Echo State Networks*.



Conclusão

As arboviroses representam um desafio significativo para a saúde pública, tanto em escala nacional quanto global, sobretudo em países tropicais como o Brasil. Na cidade do Recife, a dengue, a zika e a *chikungunya* são as enfermidades mais frequentes entre as arboviroses, já que seu cenário urbano e clima favorável para proliferação de seu principal vetor de transmissão, torna-a uma região propícia para a disseminação dessas doenças.

Considerando esses aspectos, neste estudo, foi proposto a investigação de dois algoritmos de computação de reservatório, a *Extreme Learning Machine* (ELM) e a *Echo State Network* (ESN), para a predição da quantidade de casos de dengue, zika e *chikungunya* na cidade do Recife, Pernambuco. Os resultados obtidos apontaram que ambos os algoritmos demonstraram excelentes níveis de desempenho na predição de casos de arboviroses. Para ambos os algoritmos, os valores das acurácias foram acima de 80% para a maioria das configurações avaliadas, com valores de *precision* e *recall* acima de 0,8, esses limiares foram definidos empiricamente a partir da experiência de epidemiologistas da Secretaria Municipal de Saúde do Recife como bons. Isso indica que os modelos se mostraram robustos e capazes de captar os padrões de casos de arboviroses de forma eficaz e com um baixo custo computacional. Sendo assim, mostraram um grande potencial para prever a incidência de arboviroses com



antecedência, portanto boas ferramentas de apoio para gestores de saúde, na implementação de medidas preventivas de controle de forma mais eficiente. A alta capacidade preditiva dos modelos os posiciona como importantes recursos no enfrentamento das doenças na região, possibilitando o desenvolvimento de estratégias para a tomada de decisões essenciais no combate às arboviroses. A identificação de áreas com maior incidência e a categorização dos bairros conforme o número de casos são etapas cruciais para mitigar o impacto dessas doenças na saúde pública. Esses avanços na modelagem preditiva não apenas ampliam a compreensão sobre a propagação das arboviroses, mas também permitem que as autoridades de saúde identifiquem áreas com maior quantidade de casos com antecedência. Além de otimizar recursos sabendo qual área da cidade mais precisa receber investimento e também em desenvolver um bom planejamento a longo prazo para o controle e prevenção.

Agradecimentos

O presente trabalho foi realizado com financiamento parcial da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001, da Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE), do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e ao Departamento de Ciência e Tecnologia do Ministério da Saúde (DECIT/MS). Os autores agradecem à Universidade de Pernambuco (UPE) e à Universidade Federal de Pernambuco (UFPE).

Referências

1. SINARE, Akshata B.; BARKADE, Ganesh D. Dengue: A mosquito borne disease. World Journal of Biology Pharmacy and Health Sciences, v. 14, n. 2, p. 251-261, 2023.
2. PARIJA, Subhash Chandra. Arboviruses (Arthropod-borne Viruses) and Rodent-borne Viruses. In: Textbook of Microbiology and Immunology. Singapore: Springer Nature Singapore, 2023. p. 825-846.
3. VAZIRI, Samira; POUR, Siavash Hamzeh; AKRAMI-MOHAJERI, Fateme. Zika virus as an emerging arbovirus of international public health concern. Osong Public Health and Research Perspectives, v. 13, n. 5, p. 341, 2022.
4. OEHLER, Erwan et al. Increase in cases of Guillain-Barré syndrome during a Chikungunya outbreak, French Polynesia, 2014 to 2015. Eurosurveillance, v. 20, n. 48, p. 30079, 2015.



5. FARIA, Marco Túlio da Silva et al. Saúde e saneamento: uma avaliação das políticas públicas de prevenção, controle e contingência das arboviroses no Brasil. *Ciência & Saúde Coletiva*, v. 28, p. 1767-1776, 2023.
6. ESPINAL, Marcos A. et al. Emerging and reemerging Aedes-transmitted arbovirus infections in the region of the Americas: implications for health policy. *American journal of public health*, v. 109, n. 3, p. 387-392, 2019.
7. PÉREZ, Tatiana Betancur et al. Covid-19 and Dengue Coinfection: Case Report. *Japan Journal of Clinical & Medical Research*, v. 1, n. 3, p. 1-2, 2021.
8. MARDEKIAN, Stacey K. et al. Diagnostic options and challenges for dengue and chikungunya viruses. *BioMed research international*, v. 2015, 2015.
9. VALERIO, Fernanda et al. The neurological sequelae of pandemics and epidemics. *Journal of Neurology*, v. 268, n. 8, p. 2629-2655, 2021.
10. DA SILVA, Cecilia Cordeiro et al. Spatiotemporal forecasting for dengue, chikungunya fever and Zika using machine learning and artificial expert committees based on meta-heuristics. *Research on Biomedical Engineering*, v. 38, n. 2, p. 499-537, 2022.
11. LIMA, Clarisse Lins de. Modelagem e simulação da transmissão de arboviroses baseados em agentes inteligentes para predição de casos de dengue, chikungunya e zika. 2020. Dissertação de Mestrado. Universidade Federal de Pernambuco.
12. FALCO, Javert Guimarães. *Estatística aplicada*. 2016.
13. HUANG, Guang-Bin; ZHU, Qin-Yu; SIEW, Chee-Kheong. Extreme learning machine: theory and applications. *Neurocomputing*, v. 70, n. 1-3, p. 489-501, 2006.
14. JAEGER, Herbert. The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, v. 148, n. 34, p. 13, 2001.
15. HAYKIN, S.; NIE, Junhong; CURRIE, B. Neural network-based receiver for wireless communications. *Electronics Letters*, v. 35, n. 3, p. 203-205, 1999.
16. JAEGER, Herbert. Echo state network. *scholarpedia*, v. 2, n. 9, p. 2330, 2007.
17. STEINER, Peter et al. PyRCN: A toolbox for exploration and application of Reservoir Computing Networks. *Engineering Applications of Artificial Intelligence*, v. 113, p. 104964, 2022.
18. RAMOS, Jorge Luis Cavalcanti et al. Um estudo comparativo de classificadores na previsão da evasão de alunos em EAD. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. 2018. p. 1463.