

ESTIMATIVA FUTURA DE NOVOS CASOS DE SÍFILIS CONGÊNITA E ADQUIRIDA NO BRASIL ATÉ 2030 A PARTIR DE PREVISÃO DE SÉRIES TEMPORAIS

MARIA CLARA DA SILVA MAIA^{1,2}, DANIELA TEIXEIRA BORGES^{2,3}, RENATA
DOS SANTOS RABELLO^{2,4}

1 Introdução

A sífilis, uma infecção sexualmente transmissível, segue um problema de saúde pública. Sua transmissão pode ser também vertical, gerando a sífilis congênita (Araujo et al., 2026). Nesse contexto, a previsão de séries temporais é opção para a prevenção precoce de epidemias e melhor direcionamento das campanhas em saúde pública (Wang et al., 2024).

2 Objetivos

Estimar os casos de sífilis adquirida e congênita até 2030 no Brasil com os modelos autorregressivo integrado de médias móveis sazonal (SARIMA), de memória de curto-longo prazo (LSTM) e um modelo híbrido SARIMA-LSTM, arquitetado pela média aritmética entre as previsões de ambos. Esses métodos foram escolhidos pois os modelos média móvel e redes neurais artificiais são os mais utilizados para prever infecções (Lima et al., 2022).

3 Metodologia

Trata-se de um estudo ecológico realizado entre agosto de 2025 a maio de 2026 com dados disponíveis no Sistema de Informação de Agravos de Notificação. Coletaram-se os casos mensais de sífilis congênita entre janeiro de 2014 e maio de 2024 e adquirida entre janeiro de 2014 e dezembro de 2023 no Brasil. Vale ressaltar que, apesar de dados institucionais governamentais conferirem confiabilidade, geram também importante limitação: a subnotificação de casos, a qual compromete o desempenho dos modelos (Lima et al., 2022).

1 Acadêmica de Medicina, UFFS, *campus* Passo Fundo, contato: maria.maia@estudante.uffs.edu.br.

2 Grupo de Pesquisa: Inovação em Saúde Coletiva: políticas, saberes e práticas de promoção da saúde

3 Mestre em Envelhecimento Humano, UFFS, *campus* Passo Fundo, **Coorientadora**.

4 Doutora em Epidemiologia, UFFS, *campus* Passo Fundo, **Orientadora**.

Como foram coletados dados de domínio público, foi dispensada pela Comissão Nacional de Ética em Pesquisa (CONEP) a submissão ao Comitê de Ética em Pesquisa (CEP/CONEP) com Seres Humanos da UFFS, conforme a Resolução nº 510, de 07 de abril de 2016, do Conselho Nacional de Saúde.

A análise da tendência da série temporal foi feita via teste de Mann-Kendall e, da estacionaridade, pelo teste de Dickey-Fuller Aumentado. As bibliotecas do Python Statsmodels e pyMannKendall foram utilizadas para decompor a curva e aplicar os testes de Mann-Kendall, Dickey-Fuller Aumentado e Ljung-Box. Esse último, para ratificar a presença de autocorrelação nos dados.

Os modelos SARIMA e LSTM também foram implementados em Python, a partir das bibliotecas StatsForecast e Keras. O modelo SARIMA observa a influência entre os valores passados da série e os valores futuros, além de utilizar os erros anteriores para ajustar as previsões (Wang et al., 2024). Já o modelo de Memória de Curto e Longo Prazo (LSTM) é um tipo de Rede Neural Recorrente capaz de detectar a autocorrelação entre dados de uma série temporal, pois possui conexões transmissoras entre diferentes momentos no tempo (Silva, 2022).

O teste de Ljung-Box foi aplicado também ao final da execução dos modelos para detecção de autocorrelação nos erros preditivos. A validação ocorreu pelas métricas: erro percentual absoluto médio (MAPE), raiz do erro quadrático médio (RMSE) e erro absoluto médio (MAE). Essas foram calculadas por meio da biblioteca SkLearn do Python.

4 Resultados e Discussão

4.1 Sífilis Adquirida

A decomposição da série temporal de novos casos de sífilis adquirida no Brasil entre janeiro de 2014 e dezembro de 2023 mostrou aumento linear das notificações. Em 2014, foram 51.032 casos anuais, já em 2023, 247.839. Houve queda brusca durante a pandemia de COVID-19, porém ela não influenciou a tendência da curva. Afinal, a partir do segundo semestre de 2021, o crescimento de casos continuou como se fosse uma sequência direta das notificações dos primeiros meses de 2020.

Evidenciou-se a presença de sazonalidade anual, com maior concentração de casos no primeiro semestre de cada ano. O teste de Mann-Kendall confirmou tendência crescente e o

teste de Dickey-Fuller aumentado indicou ausência de estacionariedade da série ($p = 0,659$). Além disso, o teste Q de Ljung-Box aplicado à série original apresentou $p < 0,001$, confirmando a existência de autocorrelação dos dados.

Com base na função AutoARIMA da biblioteca StatsForecast, que testa automaticamente diversos ajustes do modelo SARIMA e seleciona o mais simples entre aqueles com menor erro, foi selecionada a configuração SARIMA(0,1,2)(0,0,1)[12]. Paralelamente, o LSTM foi implementado com 200 neurônios e 100 epochs de treinamento, porque ajustes menores obtiveram métricas de erro consideradas inaceitáveis e valores maiores não melhoram o desempenho a ponto de justificar seu uso. A função de ativação ReLU e otimizador Adam para o LSTM foram adotados por serem o padrão.

Devido à ausência de dados posteriores a dezembro de 2023 e ao comportamento anômalo das notificações durante a pandemia de COVID-19, o horizonte preditivo alcançado foi até dezembro de 2025. Previsões para além desse período tiveram métricas de erro menos confiáveis, indicando que os modelos não garantem precisão em suas estimativas para longos períodos, especialmente levando em conta o possível surgimento de novas variáveis exógenas. Variáveis essas cuja influência nos casos notificados ambos os modelos falharam em avaliar.

Os dois modelos apresentaram resíduos aleatórios (Ljung-Box: SARIMA $p = 0,068$; LSTM $p = 0,321$). As métricas foram próximas, com leve vantagem do SARIMA (RMSE 2.232 vs 2.422; MAE 1.891 vs 1.922; MAPE 10% para ambos). A média aritmética obteve melhores resultados em todos os índices de erro (RMSE = 1927, MAPE = 8% e MAE = 1579) e independência do ruído (Ljung-Box: $p = 0,16$). Como os limites aceitáveis de precisão consistem em diferença de até 10% em relação aos valores reais (Araujo et al., 2026), pode-se considerar o modelo híbrido SARIMA-LSTM como válido para estimar casos de sífilis adquirida no Brasil para o horizonte de 2 anos.

As previsões do SARIMA estabilizam em 19,8 mil casos mensais a partir de março de 2025, enquanto as do LSTM declinam lentamente até 18,96 mil mensais no mesmo período. A média aritmética trouxe valores intermediários entre esses patamares.

4.2 Sífilis Congênita

A decomposição da série temporal de casos de sífilis congênita de 2014 a 2024 no Brasil indica aumento progressivo entre 2014 e 2018, ultrapassando o marco de 2.400 casos mensais no pico do período. Após esse crescimento, permaneceu em estabilidade, mantendo-se

entre 1.800 e 2.400 casos por mês. Entre 2020 e 2021, ocorreu redução temporária seguida de novo crescimento. A partir de 2022, tendência de leve declínio, embora os níveis se mantenham elevados em comparação ao início da série.

O componente sazonal indicou padrão anual recorrente com aumento dos casos no primeiro semestre do ano. O teste de Dickey-Fuller aumentado indicou que a série não era estacionária ($p = 0,13$); o teste de Mann-Kendall confirmou uma tendência crescente. O teste de Ljung-Box ($p = 2,07 \times 10^{-75}$) apontou a autocorrelação dos dados.

O ajuste ótimo do modelo SARIMA foi feito da mesma forma que com a sífilis adquirida, resultando em SARIMA(0,1,1)(0,1,2)[12]. O LSTM, inicialmente, também utilizou a mesma montagem, porém, para melhor captar o comportamento da curva, precisou-se aumentar a configuração progressivamente até 600 neurônios e 120 epochs de treinamento.

Devido a ausência de dados posteriores a junho de 2024 e a limitada confiabilidade para previsões de médio prazo, o horizonte preditivo alcançado foi até maio de 2027. Dessa vez, os modelos conseguiram fazer previsões com erros dentro do intervalo aceitável de MAPE $< 10\%$ para o período pandêmico. Assim, ou as notificações de sífilis congênita foram menos influenciadas por essa variável exógena súbita ou os modelos foram mais capazes de captar como esse impacto se deu nessa série temporal em questão. De todo modo, percebe-se uma maior confiabilidade preditiva para esse tipo de enfermidade.

O SARIMA teve desempenho preditivo relativamente superior (RMSE = 295; MAPE = 11%; MAE = 236) ao LSTM (RMSE = 344; MAPE = 15%; MAE = 322). O modelo híbrido foi superior a ambos (RMSE = 136; MAPE = 5%; MAE = 119). Até maio de 2027, o modelo SARIMA-LSTM aponta estabilidade oscilatória entre 2.000 e 2.200 notificações mensais.

5 Conclusão

O modelo híbrido SARIMA-LSTM provou-se adequado para estimar casos de sífilis congênita para os próximos 3 anos e adquirida para os próximos 2 anos no Brasil. Portanto, recomenda-se a continuidade de testes com mais modelos e outros recortes geográficos regionais com contínua validação da precisão das estimativas, pois os erros são dinâmicos para cada período analisado. Assim viabiliza-se a criação de uma ferramenta de auxílio à gestão pública.

6 Referências Bibliográficas

ARAÚJO, Thalys Kennedy Azevedo de et al. Tendência e predição da sífilis congênita no Brasil no século XXI: Um estudo de séries temporais. **Revista Gaúcha de Enfermagem**, v. 47, p e20240166, 2026. DOI: <https://doi.org/10.1590/1983-1447.2026.20250112.pt>. Disponível em: <https://www.scielo.br/j/rgenf/a/YmTZQ8xLBrXpYvVQSkggzss/?lang=pt>. Acesso em: 3 set. 2026.

LIMA, Clarisse Lins de et al. Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic Review. **Frontiers in Public Health**, v. 10, 2022. DOI: <https://doi.org/10.3389/fpubh.2022.900077>. Disponível em: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2022.900077/full>. Acesso em: 3 set. 2026.

SILVA, Gabriel Vinícius Souza da. **Deep learning com modelos baseados em LSTM para previsão de casos de sífilis adquirida no Brasil**. 2022. 37 f. TCC (Graduação) – Curso de Engenharia Biomédica, Universidade Federal do Rio Grande do Norte, Natal, RN. Disponível em: https://eb.ct.ufrn.br/wp-content/uploads/2022/03/TCC_Gabriel_Silva_correcoes_finais.pdf. Acesso em: 3 set. 2026.

WANG, Zhen D. et al. Analysis and forecasting of syphilis trends in mainland China based on hybrid time series models. **Cambridge University Press of Epidemiology and Infection**, v. 152, n. 93, p. 1, 2024. DOI: <https://doi.org/10.1017/S0950268824000694>. Disponível em: <https://www.cambridge.org/core/journals/epidemiology-and-infection/article/analysis-and-forecasting-of-syphilis-trends-in-mainland-china-based-on-hybrid-time-series-models/9953AFB9C8FB3A9EAD6BBBF30818FA2A>. Acesso em: 3 set. 2026.

Palavras-chave: Infecções Bacterianas; Doenças e Anormalidades Congênitas, Hereditárias e Neonatais; Infecções Sexualmente Transmissíveis; Algoritmos de Predição; Planejamento em Saúde.

Nº de Registro no sistema Prisma: PES-2025-0279

Financiamento:

