









Previsão de novos casos de sífilis adquirida no Brasil utilizando os modelos Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA) e de Memória em Curto e Longo Prazo (LSTM)

Prediction of new cases of acquired syphilis in Brazil using the Seasonal Autoregressive Integrated Moving Average (SARIMA) and Long Short-Term Memory (LSTM) models

Maria Clara da Silva Maia¹ , Antonio Manoel Ferreira Raymundo¹ , Daiane Conceição de Araújo¹ , Kelly de Almeida Schlager¹ , Ketlin Angelin¹ , Lídhia Cainnã de Souza Araújo¹ , Daniela Teixeira Borges² , Renata dos Santos Rabello² 

1. Discente do curso de graduação em medicina, Universidade Federal da Fronteira Sul (UFFS), campus Passo Fundo, RS, Brasil. 2. Docente do curso de graduação em medicina, Universidade Federal da Fronteira Sul (UFFS), campus Passo Fundo, RS, Brasil.

Resumo

Objetivo: estimar novos casos de sífilis adquirida no Brasil, utilizando os modelos Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA), de Memória em Curto-Longo Prazo (LSTM) e a média aritmética entre eles. **Métodos:** foram analisados dados do Sistema de Informação de Agravos de Notificação referentes às notificações mensais entre 2014 e 2023. A validação utilizou o Erro Percentual Absoluto Médio (MAPE), a Raiz do Erro Quadrático Médio (RMSE) e o Erro Absoluto Médio (MAE). **Resultados:** a série apresentou tendência crescente e sazonalidade anual. Foi não estacionária (ADF $p = 0,659$) e exibiu autocorrelação (Ljung-Box $p < 0,001$). O modelo SARIMA (0,1,2)(0,0,1)[12] e LSTM foram comparados. Ambos 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%). A média aritmética obteve melhores resultados em todos os índices de erro e independência do ruído. As previsões do SARIMA estabilizam em 19,8 mil casos a partir de março de 2025, enquanto as do LSTM declinam até 18,96 mil no mesmo período. Observaram-se tendência de aumento em todos os recortes populacionais e só existiu presença de autocorrelação dos dados no recorte de região, indicando necessidade de ver o todo para compreensão das partes. **Conclusões:** ambos os modelos foram adequados para estimar casos de sífilis adquirida, sem diferença marcante de precisão, com a média aritmética aumentando ainda mais a confiabilidade das previsões.

Palavras-chave: sífilis; infecções sexualmente transmissíveis; algoritmos de predição; planejamento em saúde.

Abstract

Objective: to estimate new cases of acquired syphilis in Brazil using the Autoregressive Integrated Moving Average (SARIMA) model, the Long-Short-Term Memory (LSTM) model, and the arithmetic mean between them. **Methods:** data from the Notifiable Diseases Information System regarding monthly notifications between 2014 and 2023 were analyzed. Validation used the Mean Absolute Percentage Error (MAPE), the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). **Results:** The series showed an increasing trend and annual seasonality. It was non-stationary (ADF $p = 0.659$) and exhibited autocorrelation (Ljung-Box $p < 0.001$). The SARIMA (0,1,2)(0,0,1)[12] and LSTM models were compared. Both models presented random residuals (Ljung-Box: SARIMA $p = 0.068$; LSTM $p = 0.321$). The metrics were close, with a slight advantage for SARIMA (RMSE 2.232 vs 2.422; MAE 1.891 vs 1.922; MAPE 10%). The arithmetic mean obtained better results in all error indices and independence from noise. SARIMA predictions stabilize at 19,800 cases from March 2025 onwards, while LSTM predictions decline to 18,960 in the same period. An increasing trend was observed in all population segments, and autocorrelation of the data was only present in the regional segment, indicating the need to consider the whole to understand the parts. **Conclusions:** both models were adequate for estimating cases of acquired syphilis, with no significant difference in accuracy, and with the arithmetic mean further increased the reliability of the predictions.

Keywords: syphilis; sexually transmitted diseases; prediction algorithms; health planning.

INTRODUÇÃO

A Questões envolvendo a saúde sexual são relevantes para a qualidade de vida da população, porém não recebem a devida atenção¹. Por isso, infecções sexualmente transmissíveis (IST) seguem como um problema global de saúde pública, afetando, especialmente, a população jovem². Entre as principais IST's estão a clamídia, HIV/AIDS, HPV, gonorreia, herpes genital, sífilis, tricomoníase, hepatites virais e molusco contagioso¹.

Quanto à sífilis, trata-se de uma doença presente na humanidade há séculos. Sua história é de riqueza antropológica, sociológica e cultural³. Seu agente etiológico é a bactéria *Treponema pallidum*, transmissível no intercuro sexual (sífilis adquirida), de forma vertical de mãe para filho na gestação (sífilis congênita) ou, menos comumente, por transfusão sanguínea⁴.

Correspondente: Maria Clara da Silva Maia. Rua Capitão Araújo, n. 20, Centro, CEP 99010-121, Passo Fundo, Rio Grande do Sul, Brasil. E-mail: maria.maia@estudante.uffs.edu.br

Conflito de interesse: Os autores declaram não haver conflito de interesse

Recebido em: 11 Dez 2025; Revisado em: 20 Jan 2026; Aceito em: 23 Jan 2026

2 Previsão de novos casos de sífilis adquirida no Brasil

Em 2020, foram registrados 7,1 milhões de casos em pessoas entre 15 e 49 anos e 425 casos de sífilis congênita por 100.000 nascidos vivos. Aproximadamente 72% das gestantes realizaram rastreamento e tratamento para sífilis no pré-natal. Estima-se que, até 2030, esses números reduzam para 0,71 milhão de casos de sífilis adquirida e menos de 50 casos congênitos por 100.000 nascidos vivos, além de 95% de cobertura no rastreamento pré-natal⁵.

Apesar disso, a notificação de casos não segue um padrão homogêneo entre os países. No Brasil, ao contrário da tendência global, a sífilis tem reemergido, exigindo maior atenção epidemiológica para fortalecer as estratégias de vigilância, prevenção e controle⁶.

Nesse contexto, a previsão de séries temporais é chave para a prevenção precoce de epidemias e melhor direcionamento das campanhas em saúde pública⁷. Essa ferramenta tem como objetivo determinar a ocorrência de epidemias futuras e, assim, viabilizar campanhas de prevenção mais assertivas. Previsões precisas possibilitam decisões mais embasadas dos gestores públicos e permitem controlar surtos antes de acontecerem⁸.

Assim, o presente estudo objetiva estimar a notificação de novos casos de sífilis adquirida no Brasil para os próximos dois anos (2024 e 2025) mediante previsão de séries temporais, utilizando os modelos autorregressivo integrado de médias móveis sazonal (SARIMA) e de memória em curto-longo prazo (LSTM) e um modelo híbrido SARIMA-LSTM, arquitetado pela média aritmética entre as previsões de ambos os modelos.

MÉTODOS

Trata-se de um estudo observacional, transversal, de caráter descritivo e com abordagem quantitativa. Foi realizado na Universidade Federal da Fronteira Sul – Campus Passo Fundo, no período de março de 2025 a setembro de 2025. Foram utilizados dados secundários disponíveis no Sistema de Informação de Agravos de Notificação (SINAN), referentes às notificações de novos casos mensais de sífilis adquirida no Brasil entre janeiro de 2014 e dezembro de 2023, de acordo com o ano de notificação e município de residência.

Foram coletados dados secundários de domínio público. Por isso, 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 (CNS). Sendo dados secundários e agregados por municípios, não existe risco de identificação pessoal dos sujeitos envolvidos. Os dados coletados serão armazenados durante cinco anos e posteriormente excluídos.

Todos os registros foram incluídos na análise. Por isso, não houve cálculo e definição de tamanho da amostra. Foi realizada uma análise das variáveis sociodemográficas disponíveis no SINAN, incluindo raça, sexo, faixa etária e região de notificação,

organizadas da seguinte forma: sexo (feminino e masculino); faixa etária em anos (0-19, 20-39, 40-59, 60+); raça (branca, preta, amarela, parda, indígena); região de notificação (Norte, Nordeste, Centro-Oeste, Sul e Sudeste). Foi calculada a presença de autocorrelação, estacionalidade e tendência para cada um desses recortes.

Os dados foram extraídos automaticamente em formato de planilha, dispensando a digitação manual. A planilha foi utilizada como base de dados para os modelos SARIMA e LSTM, implementados em linguagem Python, por meio das bibliotecas StatsForecast e Keras, respectivamente. Como as previsões em modelo híbrido costumam ser mais precisas do que às de modelos unitários⁹, foi implementado, também, um modelo híbrido, arquitetado por meio da média aritmética entre os valores previstos por cada modelo individual. O código foi escrito e executado em nuvem por meio da ferramenta Google Colabs, que é gratuita.

As bibliotecas Statsmodels e pyMannKendall foram utilizadas para decompor a série temporal e aplicar os testes de Mann-Kendall, Dickey-Fuller Aumentado e Ljung-Box. Esse último foi aplicado nos resíduos e na própria série temporal como forma de ratificar a presença de autocorrelação nos dados e independência dos ruídos da previsão. As métricas de erro foram calculadas por meio da biblioteca SkLearn.

Essas bibliotecas surgiram em um contexto de desenvolvimento de versões para Python de ferramentas estatísticas já presentes na linguagem R, objetivando tornar o Python a principal linguagem para análise estatística¹⁰. Portanto, foi considerado válido a sua experimentação na presente pesquisa.

A análise de uma série temporal inicia com sua decomposição em três componentes: tendência, sazonalidade e irregularidade. A observação da tendência pode ser realizada pelo teste d e Mann-Kendall e, a estacionaridade, pelo teste de Dickey-Fuller Aumentado. A irregularidade não pode ser analisada⁷.

Entre os modelos matemáticos mais utilizados para prever epidemias, destaca-se o ARIMA (modelo autorregressivo integrado de médias móveis). No entanto, ele não capta a influência da sazonalidade. Por isso, emprega-se, também, o modelo SARIMA (ARIMA sazonal), que tem demonstrado eficácia na previsão de doenças infecciosas em curto prazo^{11,12}. Com esses modelos só funcionam em séries estacionárias, aplicam-se transformações logarítmicas e diferenças sazonais e não sazonais para adequação do modelo⁷.

O SARIMA é definido por parâmetros que determinam os componentes sazonais e não sazonais da série (p, d, q) (P, D, Q) s , e sua configuração é auxiliada pela função de autocorrelação (ACF) e autocorrelação parcial (PACF)^{11,13}. A seleção do melhor ajuste pode ser feita por critérios de informação como Akaike (AIC) e Bayesiano (BIC), que penalizam modelos excessivamente complexos e indicam a melhor configuração baseada em menor valor numérico. A estimação dos parâmetros é feita por máxima

3 Previsão de novos casos de sífilis adquirida no Brasil

verossimilhança (MLE)⁷.

As Redes Neurais Recorrentes (RNN) constituem um modelo de rede neural voltado à análise de séries temporais, pois permitem o processamento sequencial de informações ao estabelecer conexões entre diferentes instantes de tempo. Cada neurônio recebe não apenas os dados provenientes da camada anterior, mas também o resultado de sua própria ativação em etapas anteriores, o que possibilita à rede explorar a autocorrelação dos dados em seus processos decisórios¹⁴.

Apesar de sua aplicabilidade, as RNNs apresentam limitações, como maior tempo de processamento, ausência de consideração de entradas futuras e o problema do gradiente de fuga, que compromete o ajuste adequado dos pesos. Para contornar essas dificuldades, foi desenvolvida a arquitetura de Memória de Longo e Curto Prazo (LSTM), que introduz mecanismos adicionais de controle por meio de três portas: esquecimento, entrada e saída¹⁵.

A porta de esquecimento utiliza uma função sigmoide para definir quais informações devem ser descartadas ou mantidas no estado da célula. A porta de entrada, por sua vez, determina quais novos dados serão adicionados, por meio do cálculo dos

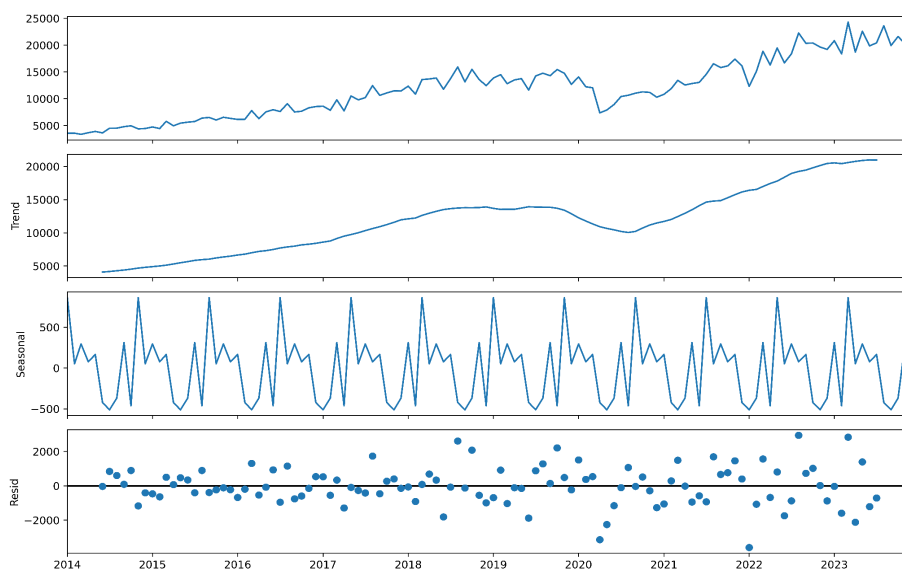
valores candidatos e da combinação com as ativações das portas correspondentes¹⁶. Após esse processo, a porta de saída define o valor final a ser transmitido, aplicando uma camada sigmoide que seleciona as informações relevantes e multiplicando-as pelos estados atualizados da célula, gerando assim a saída da rede¹⁷.

O teste de Ljung-Box é aplicado ao final da execução dos modelos para detecção de autocorrelação nos ruídos. A validação ocorre por métricas como erro percentual absoluto médio (MAPE), raiz do erro quadrático médio (RMSE) e erro absoluto médio (MAE), visto que menores valores indicam previsões mais precisas^{11,18}.

RESULTADOS

A decomposição aditiva da série temporal de novos casos de sífilis adquirida no Brasil entre janeiro de 2014 e dezembro de 2023 (figura 1) mostra aumento linear das notificações. 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 continua como se fosse uma sequência direta das notificações dos primeiros meses de 2020.

Figura 1. Decomposição da série temporal de sífilis adquirida em tendência, sazonalidade e resíduos.



Fonte: Própria, 2025

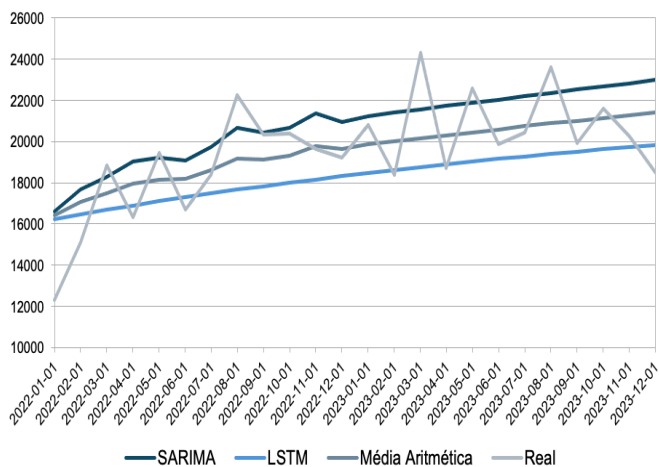
Evidenciou-se a presença de sazonalidade anual, com maior concentração de casos no primeiro semestre de cada ano. Os resíduos mostraram oscilações aleatórias de até 2.000 casos mensais, explicados por variáveis exógenas, como exemplo a pandemia de COVID-19. O teste de Mann-Kendall confirmou tendência 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.

O modelo SARIMA(0,1,2)(0,0,1)[12] foi selecionado como

a configuração ótima para a série por meio da função AutoARIMA da biblioteca StatsForecast. A rede neural LSTM foi implementada com 200 neurônios, função de ativação ReLU, otimizador Adam e 100 epochs de treinamento. Comparando as previsões do número de casos de sífilis adquirida entre 2022 e 2023 com os valores reais notificados no período, observou-se que ambos os modelos acompanharam a tendência crescente da série. O SARIMA tendeu a superestimar os casos, enquanto o LSTM demonstrou comportamento oposto, sobretudo em períodos de maior elevação. A média aritmética entre eles foi útil para compensar essas tendências. (figura 2).

4 Previsão de novos casos de sífilis adquirida no Brasil

Figura 2. Comparação de previsões verificáveis dos modelos SARIMA, LTSM e a média aritmética entre eles



Fonte: Própria, 2025

Os resíduos de ambos os modelos não apresentaram autocorrelação significativa, conforme indicado pelo teste de Ljung-Box (SARIMA $p = 0,068$; LSTM $p = 0,321$), o que sugere adequação dos modelos à série temporal. As métricas de desempenho demonstraram resultados próximos entre os dois métodos, com leve superioridade do SARIMA em relação ao LSTM. Os valores obtidos foram RMSE de 2232 para o SARIMA e 2422 para o LSTM; MAE de 1891 para o SARIMA e 1922 para o LSTM; e MAPE de 10% para ambos os modelos. Já a média aritmética entre os modelos revelou todas as métricas com melhores índices (RMSE = 1927, MAPE = 8% e MAE = 1579) e também independência na distribuição dos ruídos pelo teste de Ljung-Box ($p = 0,16$). Essa divergência de padrões evidencia a utilidade de ambos os métodos, mas reforça a necessidade de cautela na extrapolação das projeções para horizontes mais longos.

A tabela 1 e a figura 3 apresentam as previsões realizadas pelos modelos SARIMA, LSTM e a média aritmética entre eles para o período de janeiro de 2024 a dezembro de 2025. Observa-se que, enquanto o modelo SARIMA mantém estimativas relativamente estáveis a partir de março de 2025, convergindo para um valor fixo de aproximadamente 19,7 mil casos, o LSTM demonstra maior variabilidade, com projeções gradualmente decrescentes até o final de 2025.

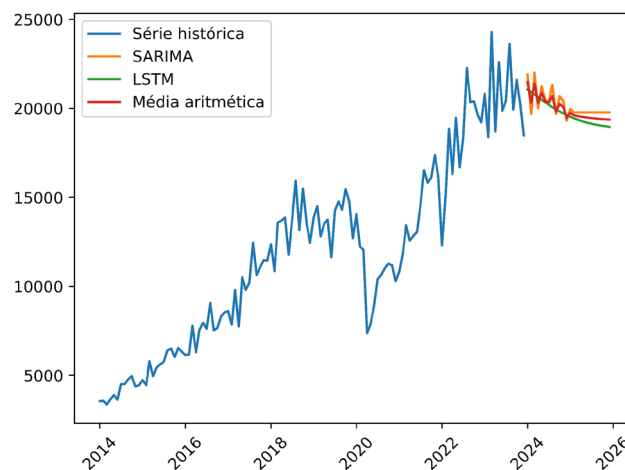
Tabela 1. Projeções de casos mensais de sífilis adquirida no Brasil de janeiro de 2024 a dezembro de 2025 pelos modelos SARIMA, LSTM e a média aritmética de suas previsões

Data	SARIMA	LSTM	Média
Jan/2024	21903	21058	21481
Fev/2024	19687	20926	20307
Mar/2024	22011	20782	21397
Abr/2024	20001	20634	20318
Mai/2024	21247	20485	20866
Jun/2024	20439	20340	20390

Data	SARIMA	LSTM	Média
Jul/2024	20438	20201	20320
Ago/2024	21323	20068	20696
Set/2024	19686	19943	19815
Out/2024	20676	19827	20252
Nov/2024	20431	19719	20075
Dez/2024	19304	19620	19462
Jan/2025	19965	19529	19747
Fev/2025	19775	19446	19611
Mar/2025	19775	19371	19573
Abr/2025	19775	19303	19539
Mai/2025	19775	19242	19509
Jun/2025	19775	19187	19481
Jul/2025	19775	19138	19457
Ago/2025	19775	19094	19435
Set/2025	19775	19055	19415
Out/2025	19775	19020	19398
Nov/2025	19775	18989	19382
Dez/2025	19775	18961	19368

Fonte: Própria, 2025

Figura 3. Projeções de casos mensais de sífilis adquirida no Brasil de janeiro de 2014 a novembro de 2025 pelos modelos SARIMA, LSTM e a média aritmética de suas previsões



Fonte: Própria, 2025

A observação dos recortes populacionais mostrou que 6,4% dos casos foram notificados na região Norte; 14,5% na região Nordeste; 49,6% na região Sudeste; 22,2% na região Sul e 7,3% na região Centro-Oeste. Homens representaram 60,7% das notificações e mulheres 39,2%, os dados restantes foram preenchidos sem o sexo. A raça branca foi 36%; a preta 10,6%; a amarela 0,94%; a parda 37,9% e a indígena 0,49%. O restante foi preenchido como raça ignorada. Quanto à faixa etária, 9,8% das notificações foram em menores de 20 anos; 58,2% entre 20 a 39 anos; 23,2% entre 40 a 59 anos e 8,3% em maiores de 60 anos. As demais notificações vieram com idade ignorada.

5 Previsão de novos casos de sífilis adquirida no Brasil

No recorte por raça/cor, observou-se que, entre pessoas pretas, a série foi classificada como não estacionária ($p = 0,61$ no teste de Dickey-Fuller aumentado), apresentando tendência crescente pelo teste de Mann-Kendall e ausência de autocorrelação significativa ($p = 0,35$ no teste de Ljung-Box). Para a população branca, verificou-se estacionariedade ($p = 0,005$), com tendência crescente e ausência de autocorrelação ($p = 0,45$). Entre os indígenas, a série foi não estacionária ($p = 0,43$), com tendência crescente e ausência de autocorrelação ($p = 0,49$). Nos pardos, também se identificou não estacionariedade ($p = 0,67$), tendência crescente e ausência de autocorrelação ($p = 0,30$). Na categoria amarela, observou-se tendência crescente e estacionariedade ($p = 0,03$), além disso, não foi identificada autocorrelação ($p = 0,14$).

No recorte por gênero, tanto em homens ($p = 0,32$) quanto em mulheres ($p = 0,27$), as séries foram classificadas como não estacionárias, com tendência crescente e ausência de autocorrelação ($p = 0,30$ e $p = 0,50$, respectivamente).

No recorte por idade, os indivíduos até 20 anos apresentaram uma série não estacionária ($p = 0,15$), com tendência de crescimento e ausência de autocorrelação ($p = 0,44$). Na faixa de 20 a 39 anos, o $p = 0,35$ confirmou a não estacionariedade, com tendência crescente e ausência de autocorrelação ($p = 0,28$). Entre pessoas de 40 a 59 anos, identificou-se não estacionariedade ($p = 0,32$), tendência crescente e ausência de autocorrelação ($p = 0,52$). Por fim, no grupo de 60 anos ou mais, também se observou não estacionariedade ($p = 0,23$), tendência crescente e ausência de autocorrelação significativa ($p = 0,57$).

No recorte por região, na Norte, observou-se não estacionariedade ($p = 0,71$), tendência crescente e autocorrelação ($p = 1,49e-105$). No Nordeste, os resultados foram semelhantes, com $p = 0,60$ (não estacionária), tendência crescente e forte autocorrelação ($p = 3,60e-98$). No Sudeste, identificou-se não estacionariedade ($p = 0,85$), tendência crescente e autocorrelação ($p = 1,77e-94$). No Centro-Oeste, a série foi não estacionária ($p = 0,71$), com tendência crescente e autocorrelação significativa ($p = 2,30e-101$). Por fim, no Sul, a série apresentou não estacionariedade ($p = 0,48$), tendência crescente e autocorrelação ($p = 1,56e-97$).

Em conjunto, os achados demonstram que as notificações de sífilis possuem autocorrelação dentro de um território, mas não entre recortes populacionais. Isso pode indicar contaminação cruzada entre os recortes, exigindo uma análise do todo para compreensão das partes. Os dados também demonstram que a doença avança no país como um todo e em todos os recortes populacionais, o que reforça a aplicabilidade de modelos de previsão, como SARIMA e LSTM para guiar políticas públicas.

DISCUSSÃO

Foram analisadas as curvas das notificações de sífilis primária, secundária e congênita na Coreia do Sul. A autocorrelação foi positiva para sífilis primária e secundária. No recorte por gênero, autocorrelação foi positiva para em homens e negativa,

em mulheres. O componente sazonal da série temporal de sífilis secundária revelou que a doença ocorre principalmente no verão¹³. Nosso estudo também obteve autocorrelação positiva, porém apenas para a população em geral. Isso pode indicar uma dinâmica de contágio diferente na Coreia do Sul entre os homens. Ademais, o aumento da incidência no verão foi ao encontro dos nossos achados.

O modelo SARIMA foi utilizado para prever casos de sífilis adquirida em homens no México. A série temporal foi não estacionária com tendência crescente de 2010 a 2015 e mudança de tendência a partir de 2016. A independência na distribuição dos ruídos pelo Teste de Ljung-Box ($p = 0,76$) foi utilizada para confirmar a viabilidade do modelo preditivo¹⁹. Esses resultados apontam que a característica de não estacionariedade está presente nas curvas da sífilis adquirida mesmo em outras localidades, e reiteram que o SARIMA é um modelo eficaz para prever seus pontos futuros.

Utilizando o modelo LSTM para prever os casos de sífilis adquirida no Brasil para o dia seguinte, obteve-se uma previsão média de 317,63 casos e um RMSE médio de 86,97 casos, representando 27,7% da quantidade média do número de casos. Ou seja, trata-se de um baixo erro. Corroborando isso, a correlação média entre o real e o previsto foi de 0,90. Quando os dados diários foram acumulados em semanas, o erro do modelo aumentou²⁰. Apesar da difícil comparação entre os resultados pelas diferenças entre horizonte preditivo e métricas de erro, é possível observar que o LSTM tem sido a escolha para testar a modelagem da curva de sífilis adquirida no Brasil.

Em comparação aos modelos SARIMA, LSTM e ao modelo híbrido SARIMA-LSTM, o modelo híbrido SARIMA-NARX apresentou as menores métricas de erro para prever casos de sífilis na China. O MAPE reduziu em 73,7%, 23,46% e 13,06%, respectivamente; o RMSE em 68,02%, 26,68% e 23,78%, respectivamente e o MAE em 70,90%, 23,00% e 21,80%, respectivamente. Achados esses que indicam que métodos híbridos preveem com maior precisão⁴¹. Tais resultados também foram encontrados em nosso estudo, visto que a média aritmética preservou a aleatoriedade do erro e reduziu os valores dessas mesmas três métricas.

Foi utilizado um modelo de rede baseado em bordas para prever a disseminação da sífilis na região de Kingston, Frontenac e Lennox & Addington, no sudeste de Ontário, Canadá. Comparando os resultados de um modelo de suscetibilidade-infecção-recuperação (SIR) baseado em rede com os de um modelo SIR tradicional de ação em massa, descobriu-se que o modelo de rede produz uma estimativa muito menor do tamanho final da epidemia²¹. Isso indica que o campo da modelagem para a compreensão da epidemiologia de infecções como a sífilis tem muitas aplicações, mas também limitações.

Um estudo previu as taxas de internação por complicações da sífilis em adultos no estado da Bahia utilizando o modelo SARIMA. A série foi decomposta pelo método aditivo, sendo possível constatar uma tendência de queda a partir de 2001

6 Previsão de novos casos de sífilis adquirida no Brasil

até 2010 e, de crescimento, entre 2012 e 2018, seguida por decréscimo a partir de 2018. Foi percebida também sazonalidade. O modelo teve aleatoriedade dos resíduos e previu permanência na redução das internações em 2021²². Esse estudo reforça a boa adequação do SARIMA para a modelagem da sífilis encontrada em nosso estudo.

Uma análise global com modelos de previsão bayesianos de idade-período-coorte até 2035 revelou aumento na incidência de sífilis para a faixa etária até 5 anos; declínio seguido de aumento na faixa etária de 5 a 19 anos; estabilidade nas faixas etárias de 20 a 29 e 75 a 79 anos; pequeno declínio para a faixas etárias maior que 80 anos; tendência ascendente para as faixas etárias de 40 a 44 e 55 a 59 anos; a faixa etária de 45 a 49 anos demonstrou uma diminuição no início, seguida por um aumento gradual; aumento inicial, seguido de declínio e, em seguida, de novo aumento para a faixa etária de 50 a 54 anos e aumento inicial, seguido de diminuição para a faixa etária de 60 a 64²³. Tais resultados pressupõem uma autocorrelação dos dados nos subgrupos de faixa etárias, o que não foi encontrado em nossa análise. Pode-se pensar que isso se deve a diferenças no comportamento da curva em cada localidade.

Uma análise dos casos mensais das 10 doenças infecciosas de notificação obrigatória mais comuns na China utilizou o ARIMA para prever a sífilis e anunciou aumento de 17% nas notificações em 5 anos²⁴. Nossos resultados apontam para uma estabilidade após um aumento histórico, indicando tendências levemente diferentes para 2026 entre Brasil e China.

O SARIMA foi utilizado para prever os casos de infecção por *Chlamydia trachomatis* na cidade de Shenzhen na China. O MAPE foi de 8,08%¹². O MAPE do SARIMA, no presente estudo, foi muito próximo ao MAPE do estudo chinês, apontando para uma possível constância na eficácia do SARIMA para estimar casos futuros de infecções sexualmente transmissíveis.

Foram utilizados os modelos ARIMA, ERNN, ARIMA-ERNN e

LSTM para prever as notificações de AIDS, gonorreia e sífilis na China. Para a previsão de 1 ano, os MAPEs para sífilis foram 9,80%, 9,55%, 8,67% e 5,79%, respectivamente. Para a previsão de 5 anos, 21,88%, 24,00%, 20,18% e 11,20%, respectivamente. Em geral, a classificação de desempenho descendente foi LSTM, ARIMA-ERNN, ERNN e ARIMA⁷. Nosso estudo também apontou superioridade do LSTM, porém muito tênue e sem grandes implicações no MAPE.

Uma análise das notificações dos casos de tuberculose no estado de Minas Gerais identificou que a série não era estacionária. Ela não apresentou sazonalidade determinística, porém apresentou sazonalidade estocástica. Foi ajustado o modelo SARIMA(1,1,0)(1,0,0)[12] para as previsões, o qual obteve boa adequação de acordo com o teste de Ljung-Box²⁵. Isso demonstra que a viabilidade do SARIMA enquanto ferramenta preditiva de doença infecciosas permanece mesmo diante de patologias com disseminações tão diferentes.

CONCLUSÃO

Os modelos preditivos são usados e validados em todo o mundo na previsão de casos de doenças infecciosas, como a sífilis adquirida. Os resultados deste estudo reiteram que eles podem ser utilizados, desde que observando o horizonte preditivo de forma crítica. As previsões indicam leve queda seguida de estabilidade próxima aos 19 mil casos mensais de sífilis adquirida para 2024 e 2025 segundo os dois modelos utilizados. A tendência crescente foi observada em todos os recortes populacionais, reforçando a necessidade de políticas públicas para a contenção da doença.

AGRADECIMENTOS

O primeiro autor recebeu uma bolsa de Iniciação Científica por essa pesquisa da Universidade Federal da Fronteira Sul (UFFS).

REFERÊNCIAS

1. Rozeira CH, Silva MF, Carneiro PE, Conzatti LA, Bolwer MB, Moura BL, et al. Principais distúrbios e infecções sexuais que afetam homens e mulheres. *Braz J Implantol Health Sci.* 2024; 6(4): 357-400. doi:10.36557/2674-8169.2024v6n4p357-400.
2. Holzmann AP, Camarinho AP, Cunha-Oliveira A, Grandi JL, Dias CL, Soares JA, et al. Fatores associados à disposição de jovens do ensino superior para adesão ao rastreio da sífilis. *SciELO Preprints.* 2024. doi:10.1590/SciELOPreprints.9513.
3. Passos MRL, Eleutério J Júnior, Bazzo ML, Carvalho RS, Nascimento AG, Oliveira MS Júnior. Syphilis, history, science, and arts: syphilis history calendar [Internet]. *Braz J Sex Transm Dis.* 2021 [acesso 2025 Jan 5]; 33. Disponível em: <https://bjstd.org/revista/article/view/1136>.
4. Pereira VR, Santos MC, Ferreira JA, Nascimento NC, Costa DL, Carvalho DS, et al. Aspectos epidemiológicos da sífilis no Brasil. *Estud Avanç Saúde Natureza* [Internet]. 2023 Jan [acesso 2024 Dez 18]; 3. Disponível: <https://www.periodicojs.com.br/index.php/easn/article/view/1111>.
5. Organização Mundial da Saúde. Setor de estratégias globais de saúde sobre HIV, hepatite viral e infecções sexualmente transmissíveis 2022-2030. Geneva: WHO; 2022 [acesso 2024 Dez 22]. Disponível em: <https://iris.who.int/bitstream/handle/10665/360348/9789240053779-eng.pdf>.
6. Ministério da Saúde [BR]. Protocolo clínico e diretrizes terapêuticas para atenção integral às pessoas com infecções sexualmente transmissíveis – IST. Brasília: Ministério da Saúde; 2022 [acesso 2024 Dez 18]. Disponível: https://www.gov.br/aid/pt-br/central-de-conteudo/pcdts/2022/ist/pcdt-ist-2022_isbn-1.pdf.
7. Zhu Z, Zhu X, Zhan Y, Gu L, Chen L, Li X. Development and comparison of predictive models for sexually transmitted diseases—AIDS, gonorrhoea, and syphilis in China, 2011–2021. *Front Public Health.* 2022 Aug; 10: 966813. doi:10.3389/fpubh.2022.966813.
8. Biggerstaff M, Johansson M, Alper D, Brooks L C, Chakraborty P, Farrow DC, et al. Results from the second year of a collaborative effort to forecast influenza seasons in the United States. *Epidemics.* 2018 Sep; 26-33. doi:10.1016/j.epidem.2018.02.003.
9. Chen L. Application progress of ensemble forecast technology in influenza

7 Previsão de novos casos de sífilis adquirida no Brasil

forecast based on infectious disease model. *Front Public Health*. 2023 Dec; 15(11): 1335499. doi:10.3389/fpubh.2023.1335499.

10. McKinney, Perktold J, Seabold S. Time series analysis in Python with statsmodels. *Proc SciPy*. 2011 Jul. doi:10.25080/Majora-ebaa42b7-012.

11. Wang ZD, Yang CX, Zhang SK, Wang YB, Xu Z, Feng ZJ. Analysis and forecasting of syphilis trends in mainland China based on hybrid time series models. *Epidemiol Infect*. 2024 May; 152: e93. doi:10.1017/S0950268824000694.

12. Weng RX, Fu HL, Zhang CL, Ye JB, Hong FC, Chen XS, et al. Time series analysis and forecasting of chlamydia trachomatis incidence using surveillance data from 2008 to 2019 in Shenzhen, China. *Epidemiol Infect*. 2020 Mar; 148: (76). doi:10.1017/S0950268820000680.

13. Song YJ, Choi H. Seasonality of syphilis in males through the 2011 to 2019 mandatory surveillance period: a cross-sectional study in South Korea. *Medicine (Baltimore)*. 2023 Dec; 102(50): e36723. doi:10.1097/MD.00000000000036723.

14. Orssatto AJ, Rizzi CB, Rizzi RL, Brun A. Aplicação de séries temporais na estimação do número de casos de dengue em Cascavel-PR. *Rev Bras Comput Apl*. 2022; 14(3): 37-50. doi:10.5335/rbca.v14i3.13483.

15. Campanhol AMC. Explicabilidade de modelo Long Short-Term Memory (LSTM) para previsão de dengue em Porto Alegre [Tese] [Internet]. Santa Maria: Universidade Federal de Santa Maria; 2024 [acesso 2025 May 3]. Disponível em: <https://repositorio.ufsm.br/handle/1/32817>.

16. Khaira U, Utomo PE, Aryani R, Weini I. A comparison of SARIMA and LSTM in forecasting dengue hemorrhagic fever incidence in Jambi, Indonesia. *J Phys Conf Ser*. 2019; 1566(1): 012054. doi:10.1088/1742-6596/1566/1/012054.

17. Othman M, Indawati R, Suleiman AA, Qomaruddin MB, Sökkalingam R. Model forecasting development for dengue fever incidence in Surabaya City using time series analysis. *Processes*. 2022; 10(11): 2454. doi:10.3390/pr10112454.

18. Tang X, Chen W, Tang SQ, Zhao PZ, Ling L, Wang C. The evaluation of preventive and control measures on congenital syphilis in Guangdong Province, China: a time series modeling study. *Infection*. 2022 Oct; 50: 1179-1190. doi:10.1007/s15010-022-01791-1.

19. Ibáñez-Cervantes G, León-García G, Vargas-De-León C, Vastro-Escarpulli G, Bandala C, et al. Epidemiological behavior and current forecast of syphilis in Mexico: increase in male population. *Public Health*. 2020 Aug; 185: 386-93. doi:10.1016/j.puhe.2020.05.057.

20. Silva GVS. Deep learning com modelos baseados em LSTM para previsão de casos de sífilis adquirida no Brasil [Tese] [Internet]. Natal: Universidade Federal do Rio Grande do Norte; 2022 [acesso 2025 Jan 5]. Disponível em: https://eb.ct.ufrn.br/wp-content/uploads/2022/03/TCC_Gabriel_Silva_correcoes_finais.pdf.

21. Zhao S, Saeed S, Carter M, Stoner B, Hoover M, Guan H, Magpantay FM. Edge-based modelling for disease transmission on random graphs: an application to mitigate a syphilis outbreak. *R Soc Open Sci*. 2025; Apr; 12(4): 250032. doi:10.1098/rsos.250032.

22. Souza IG, Jesus SJ, Araújo EM, Silva CA, Guedes EF, Feitas KS, et al. Análise temporal das internações por complicações da sífilis na Bahia, no período de 2000 a 2020. *Rev Eletr Acervo Saude*. 2024; 24(6): e14083. doi:10.25248/reas.e14083.2024.

23. Yu W, You X, Luo W. Global, regional, and national burden of syphilis, 1990-2021 and predictions by Bayesian age-period-cohort analysis: a systematic analysis for the global burden of disease study 2021. *Front Med*. 2024 Aug; 11: 1448841. doi:10.3389/fmed.2024.1448841.

24. Dai Z, Lau EHY. Characterizing and forecasting common infectious diseases in China: a time series analysis. *J Public Health Emerg*. 2021; 5(30): 1. doi:10.21037/jphe-21-ab030.

25. Fernandes FA, Vilas Bôas IA, Fernandes TJ. *J. Health Biol Sci*. 2021; 9(1): 1-6. doi: 10.12662/2317-3076jhbs.v9i1.3391.p1-6.2021.

Como citar este artigo/ How to cite this article:

Maia MCS, Raymundo AMF, Araújo DC, Schlager KA, Andelin K, Araújo LCS, et al. Previsão de novos casos de sífilis adquirida no Brasil utilizando os modelos Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA) e de Memória em Curto e Longo Prazo (LSTM). *J Health Biol Sci*. 2026; 14(1): e6208