Revista Eletrônica
Acervo Saúde
Electronic Journal Collection Health ISSN 2178-2091

O futuro da sífilis congênita no sul do Brasil: previsões com modelos matemáticos e de inteligência artificial

The future of congenital syphilis in southern Brazil: predictions using mathematical and artificial intelligence models

El futuro de la sífilis congénita en el sur de Brasil: predicciones mediante modelos matemáticos y de inteligencia artificial

Maria Clara da Silva Maia[1], Gabriela Kozak Rosin[1], Gabrielle de Jesus Lemos[1], Gustavo do Prado Dobinski[1], Lais Moretti Rodrigues[1], Vanessa Argenton Tomé[1], Vitória Londero Rech[1], Daniela Teixeira Borges[1], Renata dos Santos Rabello[1].

RESUMO

Objetivo: Estimar casos de sífilis congênita no sul brasileiro, empregando os modelos Autorregressivo Integrado de Médias Móveis Sazonal (SARIMA), Rede Neural de Memória de Curto e Longo Prazo (LSTM), suavização exponencial complexa (CES) e um modelo híbrido CES-SARIMA-LSTM. **Métodos:** Utilizaram-se dados do Sistema de Informação de Agravos de Notificação, referentes às notificações mensais registradas entre janeiro de 2014 e junho de 2024. A avaliação dos modelos foi conduzida pelo Erro Percentual Absoluto Médio (MAPE), Raiz do Erro Quadrático Médio (RMSE) e Erro Absoluto Médio (MAE), com implementação em linguagem Python. **Resultados:** Teste Dickey-Fuller Aumentado apresentou valor de p = 0,008, indicando uma série temporal estacionária; o Teste Mann-Kendall indicou uma tendência crescente e o Teste Q de Ljung-Box revelou um valor de p = 2,51e-55, confirmando autocorrelação na série temporal. O modelo SARIMA (0,1,1)(0,0,1)[12] demonstrou superioridade ao LSTM, CES e modelo triplo em todas as métricas de erro. As previsões do SARIMA apontam para redução linear relativamente estável dos casos mensais de julho de 2024 a junho de 2027, partindo de 243 até chegar em 228 casos notificados. **Conclusão:** Ações governamentais são necessárias para controlar a sífilis congênita no sul do Brasil e os modelos preditivos são ferramenta para tal.

Palavras-chave: Infecções bacterianas, Doenças e anormalidades congênitas hereditárias e neonatais, Algoritmos de predição, Planejamento em saúde.

ABSTRACT

Objective: To estimate the occurrence of new cases of congenital syphilis in southern Brazil, using the Seasonal Autoregressive Integrated Moving Average (SARIMA) model, the Long Short-Term Memory (LSTM) Neural Network, the Complex Exponential Smoothing (CES) model and a hybrid CES-SARIMA-LSTM model. **Methods:** Data from the Notifiable Diseases Information System were used, referring to monthly notifications registered between January 2014 and June 2024. The evaluation of the models was conducted using the metrics Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Mean Absolute Error

[1] Universidade Federal da Fronteira Sul (UFFS), Passo Fundo – RS.

SUBMETIDO EM: 5/2026 | **ACEITO EM: 6/2026** | **PUBLICADO EM: 8/2026**

(MAE), implemented in Python. **Results:** The Augmented Dickey-Fuller test showed a p-value of 0.008, indicating a stationary time series; the Mann-Kendall test indicated an increasing trend, and the Ljung-Box Q test revealed a p-value of 2.51e-55, confirming autocorrelation in the time series data. The SARIMA (0,1,1)(0,0,1) model [12] demonstrated superiority to the LSTM, CES, and triple model in all error metrics. SARIMA's forecasts indicate a steady relatively linear reduction in monthly cases from July 2024 to June 2027, going from 243 to 228 reported cases. **Conclusion:** Government actions are still needed to prevent an increase in cases and predictive models can be an excellent auxiliary tool for this purpose.

Keywords: Bacterial infections, Congenital hereditary and neonatal diseases and abnormalities, Prediction algorithms, Health planning.

RESUMEN

Objetivo: Estimar la ocurrencia de sífilis congénita en el sur de Brasil, utilizando los modelos promedio móvil integrado autorregresivo estacional (SARIMA), memoria a corto-largo plazo (LSTM), suavizado exponencial complejo (CES) y un modelo híbrido CES-SARIMA-LSTM. **Métodos:** Se utilizaron datos del Sistema de Información de Enfermedades de Notificación Obligatoria, refiriéndose a las notificaciones registradas entre enero de 2014 y junio de 2024. La evaluación se realizó utilizando error porcentual absoluto medio (MAPE), error cuadrático medio de la raíz (RMSE) y error absoluto medio (MAE), implementados en Python. **Resultados:** La prueba de Dickey-Fuller aumentada mostró un valor p de 0,008 (serie temporal estacionaria); La prueba de Mann-Kendall indicó una tendencia creciente y la prueba Q de Ljung-Box reveló un valor p de 2,51e-55 (autocorrelación en los datos). El modelo SARIMA (0,1,1)(0,0,1)[12] demostró superioridad sobre el LSTM, CES y el modelo triple en todas las métricas de error. Las previsiones de SARIMA indican una reducción lineal relativamente constante de los casos mensuales desde julio de 2024 hasta junio de 2027, pasando de 243 a 228 casos notificados. **Conclusión:** Aún se necesitan acciones gubernamentales para prevenir un aumento en los casos y los modelos predictivos pueden ser una excelente herramienta auxiliar para este propósito.

Palabras clave: Infecciones bacterianas, Enfermedades y anomalías neonatales congénitas y hereditarias, Algoritmos de predicción, Planificación en salud.

INTRODUÇÃO

A sífilis congênita resulta da disseminação hematogênica do Treponema pallidum da gestante infectada sem tratamento ou inadequadamente tratada para o feto via transplacentária. Nesse contexto, a transmissão vertical pode ocorrer em qualquer estágio gestacional, bem como em qualquer fase da infecção materna. O estágio da sífilis na mãe e o tempo de exposição fetal intrauterina são fatores determinantes para a probabilidade dessa transmissão (SALOMÃO R, 2023).

No que se refere às manifestações clínicas, a infecção pode apresentar diferentes desfechos. Abortamento, prematuridade ou óbito neonatal são resultados possíveis durante a gestação. Além disso, o neonato pode apresentar lesões cutâneas nas palmas das mãos, plantas dos pés, além de nas regiões perioral e perianal. A infecção pode, ainda, manifestar-se de forma tardia, permanecendo latente e apresentando manifestações clínicas apenas na infância ou na vida adulta (BERLATTO KC, et al., 2025).

Embora haja indícios de redução em alguns indicadores epidemiológicos nos últimos anos, a sífilis congênita permanece como um importante problema de saúde pública no Brasil, mantendo a taxa de incidência ainda em patamar elevado. Em 2024, o país registrou aproximadamente 9,6 casos de sífilis congênita por 1.000 nascidos vivos. Esse valor excede significativamente o parâmetro de eliminação estabelecido internacionalmente, que define como meta a ocorrência de ≤ 0,5 caso por 1.000 nascidos vivos, conforme recomendado pela Organização Pan-Americana da Saúde e pela Organização Mundial da Saúde (BRASIL, 2026).

Esse parâmetro integra os compromissos regionais assumidos pelos países para a eliminação da transmissão vertical da sífilis como problema de saúde pública. Dessa forma, apesar dos avanços na ampliação da testagem no pré-natal e na oferta de tratamento pelo Sistema Único de Saúde (SUS), a magnitude atual dos indicadores demonstra que a sífilis congênita ainda exige fortalecimento das ações de vigilância, diagnóstico oportuno e tratamento adequado das gestantes e de seus parceiros (COOPER JM, et al., 2016).

Diante da perspectiva de crescimento nos casos de sífilis congênita em território nacional, cria-se um alerta de risco iminente de sobrecarga dos serviços especializados de saúde. Para mitigar esse cenário, é fundamental priorizar o fortalecimento da Atenção Primária, investindo em várias frentes para prevenção em saúde (RAIMUNDO DML, et al., 2025).

Portanto, é essencial o acompanhamento epidemiológico da sífilis congênita nas regiões brasileiras por meio de modelos preditivos. Esse instrumento identifica padrões de comportamento para as doenças e propõe-se a fazer previsões sobre o modo como elas irão se comportar. Tal mecanismo é interessante no contexto da gestão de saúde pública, tendo em vista que permite a alocação inteligente dos recursos humanos e financeiros, a avaliação de intervenções propostas e sua efetividade, a identificação de fatores de risco ou de agravos e o planejamento de ações emergenciais antecipadas (OLIVEIRA E CRUZ DM, et al., 2024).

Assim, buscando aplicar essas ferramentas, o presente estudo teve como objetivo estimar os casos notificados de sífilis congênita, em um horizonte preditivo de 3 anos, na Região Sul do Brasil. Para alcançar esse objetivo, serão utilizados os modelos autorregressivo integrado de médias móveis sazonal (SARIMA), de memória de curto-longo prazo (LSTM), suavização exponencial complexa (CES) e um modelo híbrido SARIMA-LSTM-CES, arquitetado pela média aritmética entre as previsões dos três modelos.

MÉTODOS

Trata-se de um estudo ecológico realizado na Universidade Federal da Fronteira Sul – Campus Passo Fundo, no período de fevereiro de 2026 a maio de 2026. Foram utilizados dados secundários disponíveis no Sistema de Informação de Agravos de Notificação (SINAN), referentes às notificações de novos casos mensais de sífilis congênita entre janeiro de 2014 e junho de 2024 na região sul do Brasil, de acordo com o ano de notificação e município de residência.

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. 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. Não houve cálculo e definição de tamanho da amostra, pois, por incluir todos os registros, a análise caracteriza-se como um censo. Foi realizada uma análise das variáveis sociodemográficas disponíveis no SINAN, incluindo raça, sexo, faixa etária materna e região de notificação, organizadas da seguinte forma: sexo (feminino e masculino); faixa etária materna em anos (0-19, 20-39, 40-59); raça (branca, preta, amarela, parda, indígena); estado de notificação (Santa Catarina, Rio Grande do Sul, Paraná).

Os dados foram extraídos em formato de planilha, dispensando a digitação manual. A planilha foi utilizada como base de dados para os modelos SARIMA, CES e LSTM, implementados em Python, a partir das bibliotecas StatsForecast e Keras. O código foi escrito e executado em nuvem por meio do Google Colabs, que é gratuito.

O modelo ARIMA é um acrônimo para modelo autorregressivo (AR) integrado (I) de média móvel (MA). As respectivas ordens de cada modelo são representadas por ARIMA (p, d, q). Ele pode ser adaptado para captar sazonalidade na série temporal, tornando-se o ARIMA sazonal (SARIMA). Nessa versão, é grafado como SARIMA (p, d, q) (P, D, Q)s. O componente “s” define o número de observações (WANG ZD, et al., 2024). O modelo AR observa a influência entre os valores passados da série com os valores futuros, o modelo MA utiliza os erros anteriores para ajustar as previsões e o modelo I diferencia a série, tornando-a estacionária (TANG X, et al., 2022).

O modelo de Memória de Curto e Longo Prazo (LSTM) é um tipo de Rede Neural Recorrente. Essa é uma categoria de rede neural capaz de detectar a autocorrelação entre dados de uma série temporal, pois possui conexões transmissoras entre diferentes momentos no tempo (ORSSATTO AJ, et al., 2022). Seus neurônios recebem dados da camada anterior e da sua própria ativação na passagem pregressa, tendo uma arquitetura de três portas: esquecimento, entrada e saída A porta de esquecimento define quais informações excluir do estado da célula. Já a de entrada, determina qual informação adicionar ao estado da célula (YAGNIK H, et al., 2026). Após a célula de memória passar pela porta de entrada e esquecimento, a porta de saída gera a previsão do próximo valor a partir do produto de uma camada sigmoide e uma função de ativação (OTHMAN M, et al., 2022).

A suavização exponencial simples utiliza uma média ponderada de dados históricos para prever valores futuros. Requer apenas um parâmetro de nível (α) entre 0 e 1 para controlar a rapidez com que a influência das observações passadas diminui. É considerado adequado apenas para dados de séries temporais curtas, sem tendência clara ou padrão sazonal (SILVA WS, et. al., 2025). Quando objetiva-se prever séries com tendência e/ou sazonalidade, recomenda-se utilizar a suavização exponencial complexa. Nesse contexto, além do nível, existem parâmetros de tendência (β) e sazonalidade (γ) (ALVES CdaC, 2019). É possível também avaliar a sazonalidade transformando o parâmetro de nível em um número complexo. Assim, a parte real (α0) atua como parâmetro de nível sem sazonalidade e, a parte imaginária (α1), como parâmetros de nível sazonal. Nesse caso, em vez de α estar entre 0 e 1, tem-se a regra de que a raiz quadrada da soma de $\alpha 0^2$ e $\alpha 1^2$ esteja entre 0 e 1 (SVETUNKOV I, et al., 2022).

As previsões em modelo híbrido costumam captar melhor as variações aleatórias na incidência de casos de doenças infecciosas e serem mais precisas do que as de modelos unitários (CHEN L, 2023). Por isso, foi implementado também um modelo híbrido, arquitetado por meio da média aritmética entre os valores previstos por cada modelo individual.

Iniciou-se o estudo da série temporal com sua decomposição em três componentes: tendência, sazonalidade e irregularidade. A análise da tendência foi via teste de Mann-Kendall e, da estacionaridade, pelo teste de Dickey-Fuller Aumentado. 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, para ratificar a presença de autocorrelação nos dados.

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

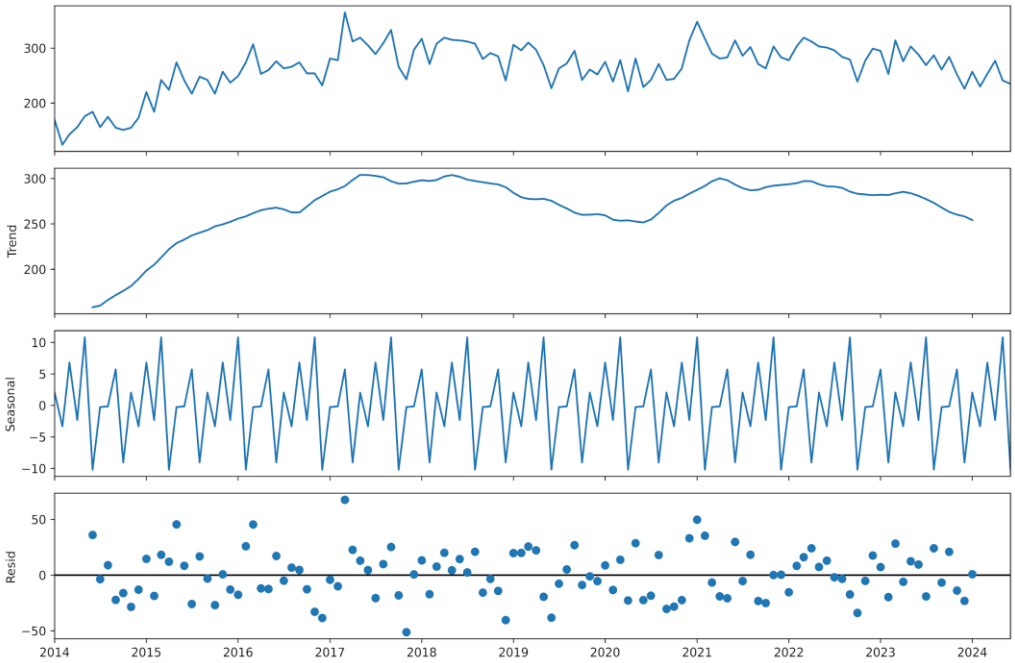
RESULTADOS

Os dados coletados mostraram que a Região Sul do Brasil registrou 40.646 casos de sífilis congênita no período analisado. Em relação à distribuição entre os estados, há concentração da maioria no Rio Grande do Sul (57,6%), seguido pelo Paraná com 25,3% e Santa Catarina com 17,1%.

Em relação ao perfil materno, observou-se um predomínio na faixa etária entre 20 e 29 anos, com 56,7%. A segunda maior faixa é de 30 à 39 anos (20,8%), seguidos de 15 a 19 anos (17,6%), 40 anos ou mais (2,5%) e, por fim, entre 10 e 14 anos (0,5%). Já no recorte de raça/cor, os dados indicam que a população autodeclarada branca contém a maior porcentagem de notificações (70%), enquanto as raças parda e preta representam 18% e 9%, respectivamente. No que tange o gênero dos recém-nascidos na série histórica de 2014 a 2023, observou-se uma distribuição bastante equilibrada, com 50,2% sendo do sexo feminino.

Inicialmente, exploraram-se os dados históricos de notificações de Sífilis Congênita na região Sul do Brasil, no período de 2014 a 2024. A decomposição dessa série temporal (**Figura 1**) revelou dinamicidade na evolução dos registros, observando-se elevações contínuas com quedas e retomadas posteriores na tendência, especialmente em 2016 e 2020, além de aparente breve redução relativa no último ano coletado. Verificou-se, ainda, uma alteração sazonal cíclica, marcada por picos de notificações seguidos de quedas abruptas. Os ruídos observados oscilam em cerca de 100 casos, que não são explicados por tendência ou sazonalidade.

Figura 1 - Decomposição da série temporal das notificações de Sífilis Congênita do Sul do Brasil de 2014 a 2024.



Fonte: Maia MCS, et al., 2026.

A seguir, aplicaram-se testes estatísticos. Assim, o Teste Dickey-Fuller Aumentado (ADF) apresentou valor de p = 0,008, indicando uma série temporal estacionária; o Teste Mann-Kendall indicou uma tendência crescente e o Teste Q de Ljung-Box revelou um valor de p = 2,51e-55, confirmando a autocorrelação nos dados na série temporal.

Quando foram feitos os mesmos testes nas séries temporais de cada um dos estados do sul, o Teste Q de Ljung-Box apontou que todas as séries possuíam autocorrelação dos dados (SC: p = 0,025; RS: p = 0,003; PR: p = 9,61e-37). Já o teste de Mann-Kendall, relatou tendência crescente no Paraná e em Santa Catarina e ausência de tendência significativa no Rio Grande do Sul. Quanto ao Teste ADF, houve comportamento estacionário da série no Rio Grande do Sul e em Santa Catarina, mas não no Paraná (SC: p = 3,34e-32; RS: p = 2,01e-38; PR: p = 9,61e-37).

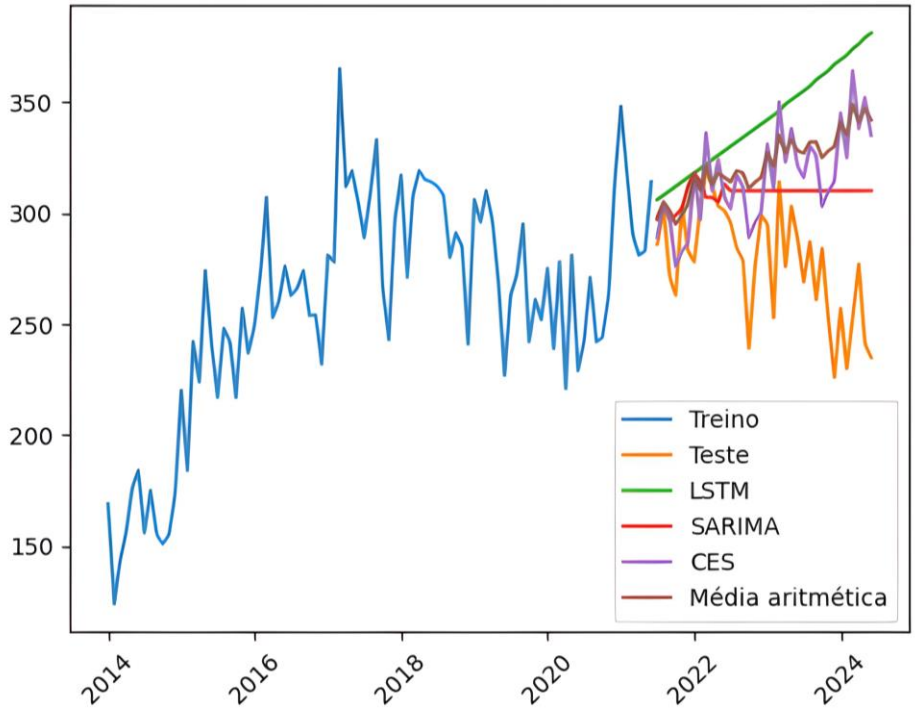
Em seguida, realizou-se o treinamento e avaliação dos modelos computacionais. Utilizaram-se dados de janeiro de 2014 a junho de 2021 como conjunto de treinamento, a partir dos quais foram realizadas previsões para o período de 2021 a 2024. Quatro abordagens de previsão foram implementadas para modelar a evolução da Sífilis Congênita: SARIMA, CES, LSTM e a média aritmética de todos. Por meio da função AutoARIMA da biblioteca StatsForecast, encontrou-se a ordem ótima para o ajuste do modelo SARIMA em (0,1,1)(0,0,1)[12]. Embora o teste ADF indicasse estacionaridade da série, foi incluída diferenciação de primeira ordem, possivelmente em razão do critério de otimização adotado e da sozinalidade da série.

Ainda com a biblioteca StatsForecast, utilizou-se a função AutoCES e os parâmetros otimizados foram: $\alpha 0 = 1,48$, $\alpha 1 = 1,01$ e $\beta = 0,02$. Novamente, observou-se um resultado inesperado, pois a raiz quadrada da soma de $\alpha 0^2$ e $\alpha 1^2$ resultou em 1,79. Isso indicou que a série temporal requeria suavização agressiva e ajuste sazonal. Apesar disso, foi considerado que o modelo CES ainda poderia ser capaz de prever, desde que tivesse um desempenho adequado nas métricas de erro. Paralelamente, o LSTM foi configurado com 400 neurônios, 200 epochs, ativação ReLU e otimizador Adams.

Aplicando-se o Teste de Q Ljung-Box nos resíduos dos modelos, obtiveram-se os seguintes valores de p: LSTM = 2,31e-15, SARIMA = 0,0003, CES = 1,003e-11 e média aritmética p = 2,38e-12. À exceção do SARIMA, todos apresentaram inadequação ao teste, indicando que não captaram bem as nuances da curva.

O desempenho foi mensurado em três métricas globais de erro: Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE) e Erro Percentual Absoluto Médio (MAPE). Quanto aos resultados, o modelo SARIMA alcançou RMSE = 40, MAPE = 0,12 e MAE = 32; o modelo LSTM apresentou RMSE = 77, MAPE = 0,25 e MAE = 64; o modelo CES obteve RMSE = 51, MAPE = 0,15 e MAE = 40; já a média aritmética apresentou RMSE = 55, MAPE = 0,17 e MAE = 44. Vale ressaltar que o SARIMA obteve menor erro em todas as métricas, o que é compatível com o fato de ter sido o único com aleatoriedade no erro pelo Teste Q Ljung-Box. A **Figura 2** apresenta a comparação entre o acerto dos modelos.

Figura 2 - Validação das previsões dos modelos comparando suas estimativas com os valores reais de junho de 2021 a maio 2024.



Fonte: Maia MCS, et al., 2026.

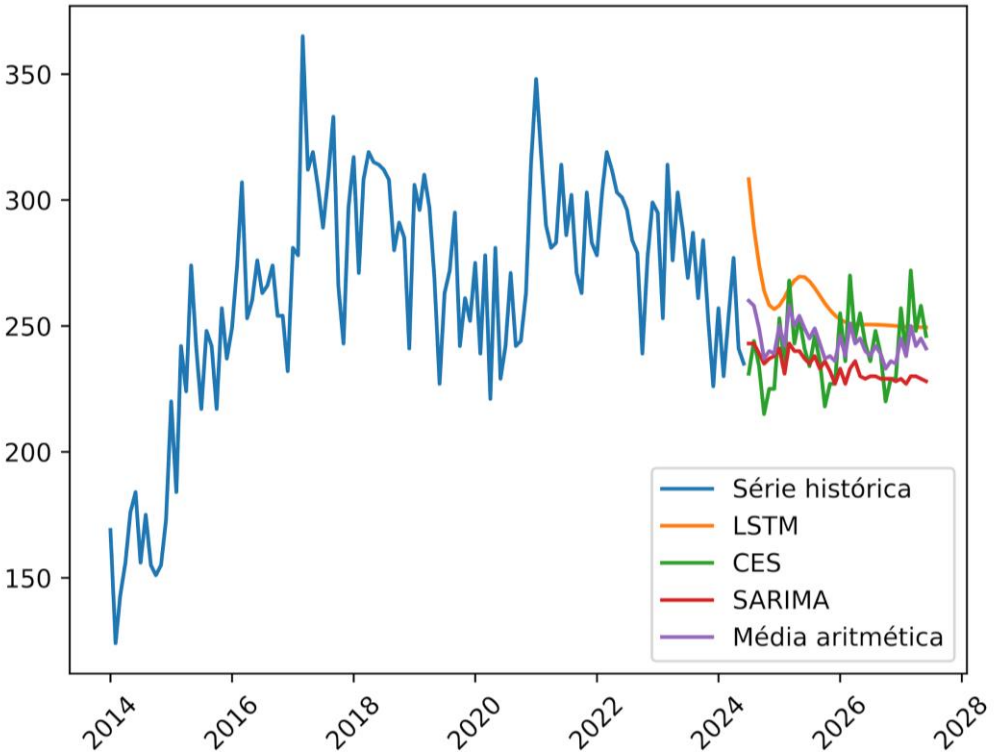
Os modelos então foram empregados para traçar um cenário futuro das notificações de Sífilis Congênita no Sul do Brasil, de julho de 2024 a junho de 2027, dados sistematizados na **Tabela 1**. As projeções obtidas pelo SARIMA sugerem que, à exceção de intervenções disruptivas nas políticas de saúde pública, o número de casos se manterá elevado, flutuando sazonalmente. A representação gráfica comparativa encontra-se na **Figura 3**.

Tabela 1 - Projeções das notificações de Sífilis Congênita até junho de 2027.

Mês/Ano	LSTM	SARIMA	CES	Média Aritmética
07/24	308	243	231	260
08/24	288	243	244	258
09/24	274	239	234	249
10/24	263	235	215	237
11/24	258	237	225	240
12/24	256	238	225	239
01/25	258	241	253	250
02/25	261	231	234	242
03/25	264	243	268	258
04/25	267	240	243	250
05/25	269	240	253	254
06/25	269	237	241	249
07/25	267	235	234	245
08/25	265	238	246	249
09/25	261	233	236	243
10/25	258	236	218	237
11/25	256	232	227	238
12/25	254	227	227	236
01/26	252	233	255	246
02/26	251	227	236	238
03/26	250	233	270	251
04/26	250	236	245	243
05/26	250	230	255	245
06/26	250	229	243	240
07/26	250	230	236	238
08/26	250	230	248	242
09/26	250	229	239	239
10/26	250	229	220	233
11/26	250	229	229	236
12/26	249	228	229	235
01/27	249	229	257	245
02/27	249	227	238	238
03/27	249	230	272	250
04/27	249	230	248	242
05/27	249	229	258	245
06/27	249	228	246	241

Fonte: Maia MCS, et al., 2026.

Figura 3 - Série histórica e projeção de notificações de Sífilis Congênita no Sul do Brasil até maio de 2027.



Fonte: Maia MCS, et al., 2026.

DISCUSSÃO

Berlatto KC, et al. (2025) afirmam que a análise epidemiológica via SINAN dos casos de sífilis congênita na Região Sul do Brasil entre 2018 e 2022 permite encontrar como grupo vulnerável as gestantes adolescentes e jovens adultas.

O presente estudo utilizou uma coleta mais ampla, incluindo o período de janeiro de 2014 a junho de 2024. Porém, ainda assim, a maior prevalência de sífilis congênita foi nas mães entre 20 e 29 anos. Este achado é corroborado por Feng S, et al. (2026), que demonstram que mulheres mais jovens têm maior probabilidade de transmitir sífilis para seus filhos do que mulheres mais velhas.

Para concluir tal coisa, Feng S, et al. (2026) construíram uma calculadora online, que integra variáveis sociais e demográficas, prevendo a chance individual de uma paciente grávida transmitir sífilis ao filho. Assim, a técnica preditiva utilizada foca em medidas de saúde individuais, diferindo-se da abordagem aqui adotada, que visa parâmetros coletivos de previsão. Nesse sentido, há várias formas de aplicações para modelos preditivos, apontando para um futuro promissor de pesquisas na área.

Utilizando uma técnica parecida, Teixeira IV, et al. (2025) utilizam diferentes modelos de inteligência artificial para prever a ocorrência de sífilis congênita no filho de uma dada gestante com base em variáveis sociodemográficas e clínicas que determinam os casos positivos. Assim, obtém-se um instrumento de identificação de grupos de risco entre a população de gestantes. Novamente, ressalta-se a diferença entre prever desfechos individuais e sociais e que ambas as abordagens em conjunto são uma boa alternativa no controle das doenças congênitas.

Outra abordagem diferenciada acerca dos métodos de modelagem foi proposta por Young SD, et al. (2018), que utilizaram dados de redes sociais para prever casos de sífilis. Foi encontrada uma relação positiva significativa entre tweets e casos de sífilis primária e secundária e sífilis latente precoce. Desse modo, percebe-se como o arsenal de ferramentas envolvendo estimativas futuras para auxiliar medidas precoces em gestão de saúde pública é vasto e deve ser explorado.

O estudo de Souza IG, et al. (2023) demonstrou que o SARIMA obteve resíduos aleatórios na previsão das internações por sífilis na Bahia. Pela decomposição da série temporal, foram detectadas tendências crescentes e decrescentes alternadas, além de sazonalidade. O horizonte testado foi para 12 meses. Os dados aqui apresentados corroboram todos esses resultados, o que pode indicar que as internações por sífilis e a transmissão vertical podem apresentar comportamentos estatísticos parecidos enquanto complicações da sífilis adquirida.

Segundo análise de Tang X, et al. (2022), a série temporal de casos de sífilis congênita na província de Guangdong, na China, se mostrou não estacionária. Um modelo ARIMA foi validado pela distribuição aleatória dos ruídos pelo Teste de Ljung-Box ($p > 0,05$). Os casos previstos para 2021 ficaram num patamar de epidemia de baixo nível. Esses resultados divergem dos aqui apresentados, pois encontrou-se uma série estacionária. Porém, reafirma-se aqui a adequação do modelo ARIMA sazonal utilizado, único entre os testados que apresentou ruído branco como erro preditivo.

A pesquisa de Araujo TKA, et al. (2026) demonstrou que a capacidade preditiva do SARIMA pode ser comprometida em cenários de ruptura estrutural, como observado durante a pandemia da COVID-19. Tal achado reforça a limitação do modelo em lidar com eventos inesperados, uma vez que sua estrutura se baseia exclusivamente em padrões históricos. Em contrapartida, no presente estudo, o SARIMA apresentou desempenho superior em relação aos demais modelos testados, inclusive em 2021, talvez por 2020 ter sido incluído no conjunto de treino. Os erros associados ao SARIMA aqui encontrados estão relacionados à dificuldade em manter a precisão com horizontes preditivos muito grandes, com a acurácia diminuindo de forma importante após 1 ano de previsões.

Raimundo DML, et al. (2025) utilizaram os modelos suavização via médias móveis e equação das tendências temporais para prever os casos futuros de sífilis congênita até 2025 no estado do Rio Grande do Norte. O perfil mais comum dos acometidos foi sexo feminino e cor parda. As previsões apontaram relativa estabilidade dos casos. Já os resultados aqui apresentados, indicam também um predomínio no sexo feminino e previsões de estabilização. Contudo, a população autodeclarada branca foi a mais acometida, o que pode ser resultado de diferentes proporções raciais nas regiões do Brasil.

Carneiro DO, et al. (2023) coletaram dados do SINAN entre 2007 e 2021 e utilizaram o modelo de suavização exponencial para modelar as notificações de sífilis congênita no Brasil até 2025. O aumento de casos foi previsto para todas as regiões, exceto a Sul, para a qual foi estimada uma queda. Essas previsões vieram a se concretizar a partir da análise dos anos seguintes feitas no presente estudo, que coletou ainda dados referentes a 2022, 2023 e 2024. Desse modo, fica evidente que os modelos preditivos são uma ferramenta com amplo potencial de uso para a gestão em saúde pública.

Aziz AA, et al. (2025) implementou um modelo híbrido de suavização exponencial e Otimização de Acasalamento de Barnacles (BMO) com o objetivo de prever casos da COVID-19 na Malásia. O modelo CES isolado obteve baixo desempenho preditivo, fato que foi explicado pela sua excessiva simplicidade, mas sua associação ao BMO melhorou muito as métricas de erro.

Esse fenômeno também foi observado nos resultados aqui apresentados, pois o CES isolado teve um desempenho pior do que quando associado ao SARIMA. Contudo, o LSTM apresentou métricas de erro ainda piores que o CES, mesmo sendo um modelo de maior complexidade. Desse modo, infere-se que os desempenhos dos modelos são muito variáveis e não é recomendado fixar um como melhor de forma absoluta.

O estudo de Kumar S, et al. (2021) demonstrou a eficácia do modelo LSTM em estimar a propagação da COVID-19. Nos resultados aqui obtidos, o LSTM apresentou desempenho inferior em comparação aos demais métodos. Essa divergência pode ser atribuída às diferenças intrínsecas entre as séries temporais, uma vez que a disseminação da COVID-19 apresenta comportamento altamente não linear e influenciado por múltiplos fatores dinâmicos, enquanto a série de sífilis congênita apresenta maior regularidade e sazonalidade, favorecendo modelos estatísticos tradicionais. Dessa forma, os achados reforçam que não há um modelo universalmente superior, mas sim abordagens mais adequadas a depender do contexto.

Echegaray F, et al (2025) usaram o modelo Holt-Winters para prever os casos de sífilis na gestação em Porto Alegre. Os resultados indicaram estabilidade e, ao serem validados, percebeu-se tendência em subestimar a realidade. É importante ressaltar a diversidade de modelos preditivos que podem e devem ser validados para prever as mais diversas doenças. Assim, pode-se descobrir meios de deixá-los cada vez mais refinados e precisos.

Conforme Pavinati G, et al. (2025), foi evidenciada estabilidade nas notificações de sífilis congênita na região Sul do Brasil entre 2014 e 2023. No presente estudo, a previsão do modelo SARIMA, cujo erro foi menor, demonstrou que esse comportamento de estabilidade continuará até junho de 2027. Cabe analisar que o Brasil já estava longe do objetivo da redução da transmissão vertical de sífilis até o ano de 2030, como alertou Pavinati G, et al. (2025). A estabilidade encontrada no modelo SARIMA afirma que essa situação persistirá. Dessa forma, ações de prevenção, educação e diagnóstico se fazem extremamente necessárias para atingir o objetivo de redução da transmissão vertical dessa doença.

Lin Y, et al. (2025) analisa a carga global de HIV e outras infecções sexualmente transmissíveis (incluindo a sífilis) em mulheres entre 15 e 49 anos. Em relação às doenças, observou-se a diminuição da incidência de HIV, em conjunto com o aumento da incidência das demais infecções. Ao utilizar o modelo ARIMA, previu-se a manutenção dessa tendência pelos próximos anos. Em comparação com o presente estudo, que utiliza o modelo ARIMA sazonal e ainda outros modelos, os dois encontram-se em consonância, evidenciando a necessidade de monitoramento contínuo da transmissão de sífilis entre a população.

CONCLUSÃO

Os modelos preditivos são usados e validados em todo o mundo na previsão de casos de doenças infecciosas, como a sífilis congênita. Nesse contexto, os resultados deste estudo corroboram a aplicabilidade dessa técnica, desde que observando o horizonte preditivo de forma crítica e analisando qual o melhor modelo para aquela série temporal específica. No caso da presente análise, observou-se melhor desempenho do SARIMA nas métricas de erro utilizadas. As previsões do SARIMA apontam para uma redução linear relativamente estável dos casos mensais de julho de 2024 a junho de 2027, partindo de 243 até chegar em 228 casos notificados. Esse nível indica uma leve queda em relação ao registrado anteriormente e um retorno às métricas encontradas durante o início da pandemia da COVID-19, em patamar ainda elevado. Portanto, reforça-se a necessidade de manutenção e fortalecimento das ações de prevenção, diagnóstico oportuno e tratamento adequado das gestantes e parceiros, além do uso de ferramentas preditivas como apoio ao planejamento em saúde pública. Assim, evita-se a manutenção de patamares elevados de casos.

AGRADECIMENTOS E FINANCIAMENTO

O primeiro autor recebeu uma bolsa de iniciação científica da Universidade Federal da Fronteira Sul (UFFS).

REFERÊNCIAS

1. ALVES CC, et al. Aplicação de métodos estatísticos com suavização exponencial dupla e tripla para previsão de demanda na gestão de estoques. Revista Produção Online: Revista Eletrônica de Engenharia de Produção e Correlatas. 2019; 19(3): 1001-26.
2. ARAUJO TKA, et al. Tendência e predição da sífilis congênita no Brasil no século XXI: Um estudo de séries temporais. Revista Gaúcha de Enfermagem, 2026; 47: e20240166.
3. AZIZ AA, et al. A hybrid simple exponential smoothing-barnacles mating optimization approach for parameter estimation: Enhancing COVID-19 forecasting in Malaysia. MethodsX, 2025; 14: 103347.
4. BERLATTO KC, et al. Sífilis Congênita na Região Sul do Brasil (2018-2022): Perfil Epidemiológico com Discussão das Possíveis Alterações Anatômicas Associadas. Brazilian Journal of Implantology and Health Sciences, 2025; 7(1): 357-368.

5. BRASIL. 2026. Sífilis congênita. Disponível em: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/s/sifilis/gestantes/congenita. Acesso em: 8 maio 2026.
6. CARNEIRO, DO, et al. Utilização do modelo de suavização exponencial para avaliação temporal da situação epidemiológica da sífilis congênita no Brasil. Diálogos & Ciência, 2023; 3(1): 19-34.
7. CHEN L. Application progress of ensemble forecast technology in influenza forecast based on infectious disease model. Frontiers in Public Health, 2023; 15(11): 1335499.
8. COOPER JM, et al. In time: the persistence of congenital syphilis in Brazil - More progress needed!. Em tempo: a persistência da sífilis congênita no Brasil - Mais avanços são necessários!. Revista paulista de pediatria, 2016; 34(3): 251–253.
9. ECHEGARAY F, et al. Repercussions of the COVID- 19 pandemic on maternal and congenital syphilis in South Brazil: a time series analysis 2010-2022. BMC Infectious Diseases, 2025; 25(528).
10. FENG S, et al. Development and validation of an individualised nomogram to predict mother-to-child transmission in pregnant women with syphilis in China: a retrospective cohort study. BMJ Open, 2024; 14(4): e080211.
11. KUMAR S, et al. Forecasting the spread of COVID-19 using LSTM network. BMC Bioinformatics, 2021; 22(6): 316.
12. LIN Y, et al. The trend analysis of HIV and other sexually transmitted infections among women of childbearing aged 15 to 49 years from 1990 to 2021 and its forecasting from 2022 to 2030. Front. Public Health, 2025; 13: 1702289.
13. OLIVEIRA E e CRUZ DM, et al. Inteligência epidemiológica, investimento em tecnologias da informação e as novas perspectivas para o uso de dados na vigilância em saúde. Cadernos de Saúde Pública, 2024; 40(8): e00160523
14. ORSSATTO AJ, et al. Aplicação de séries temporais na estimação do número de casos de dengue em Cascavel-PR. Revista Brasileira de Computação Aplicada, 2022; 14(3): 37-50.
15. OTHMAN M, et al. Model Forecasting Development for Dengue Fever Incidence in Surabaya City Using Time Series Analysis. Processes, 2022; 10(11): 2454.
16. PAVINATI G, et al. Temporal analysis of gestational and congenital syphilis indicators in Brazil: toward the elimination of vertical transmission by 2030? Revista Brasileira de Epidemiologia, 2025; 28: e250028.
17. RAIMUNDO DML, et al. Análise espaço-temporal da sífilis congênita no Rio Grande do Norte, Brasil, de 2008 a 2022. Revista de Epidemiologia e Controle de Infecção, 2025; 15(3).
18. RAIMUNDO DML, et al. Sífilis congênita: análise de tendência temporal e projeção de casos. Acta Paulista de Enfermagem, 2025; 38.
19. SALOMÃO R. Infectologia: Bases Clínicas e Tratamento. 2 ed. Guanabara Koogan: Rio de Janeiro - RJ, 2023; 77.
20. SILVA WS, et al. Seleção de um método de previsão de demanda oriundo da aquicultura na Bacia do Rio doce em Minas Gerais (MG) e Espírito Santo (ES). Revista Produção Online, 2025; 25(3): 5336.
21. SOUZA IG. Análise temporal das internações por complicações da sífilis na Bahia, no período de 2000 a 2020. Revista Eletrônica Acervo Saúde, 2024; 24(6): 1-10.
22. SVETUNKOV I, et al. Complex exponential smoothing. Naval Research Logistics, 2022; 69(8): 1108-1123.
23. TANG X, et al. The evaluation of preventive and control measures on congenital syphilis in Guangdong Province, China: a time series modeling study. Infection, 2022; 50: 1179–1190.
24. TEIXEIRA IV, et al. Predicting congenital syphilis cases: A performance evaluation of different machine learning models. PLoS ONE, 2023; 18(6): 1-25.
25. WANG ZD, et al. Analysis and forecasting of syphilis trends in mainland China based on hybrid time series models. Cambridge University Press of Epidemiology and Infection, 2024; 152(93): 1-11.
26. YAGNIK H, et al. Deep Learning for Air Pollution Prediction: A Systematic Review of CNN, LSTM, GNN, and Transformer Architectures. Arch Computat Methods Eng, 2026.
27. YOUNG SD, et al. Using social media as a tool to predict syphilis. Preventive Medicine, 2018; 109: 58-61.