

Predição de Sepses Neonatal com Aprendizado de Máquina: Validação de Protótipo a partir de Dados Estruturados e Textuais.

Luciano Ludwig Heling¹; Amanda Jhennifer Marques Vieira²; Lauren Sallaberry Ferreira Stefanello³; Leticia Lima Junqueira⁴; Mariana Vargas da Cunha⁵; Simone Pont Zambonato Macluf⁶

¹Universidade Federal de Pelotas – llhelming@inf.ufpel.edu.br

²Universidade Federal de Pelotas – ajmvieira@inf.ufpel.edu.br

³Hospital Escola UFPEL/EBSERH – lauren.sallaberry@ebserh.gov.br

⁴Hospital Escola UFPEL/EBSERH – leticia.junqueira@ebserh.gov.br

⁵Hospital Escola UFPEL/EBSERH – mariana.vargas@ebserh.gov.br

⁶Hospital Escola UFPEL/EBSERH – simone.macluf@ebserh.gov.br

1. DESCRIÇÃO DA INOVAÇÃO

A sepsis neonatal é uma infecção grave e uma das principais causas de morbidade e mortalidade em recém-nascidos no mundo, sendo responsável por uma parcela significativa dos óbitos neonatais (WHO, 2022). O diagnóstico precoce é um desafio, pois os sinais clínicos são inespecíficos e a confirmação por hemocultura, que é o padrão-ouro, pode demorar dias (WANG et al., 2021). Essa demora leva ao uso empírico e muitas vezes excessivo de antibióticos, contribuindo para o risco de resistência bacteriana (MANI et al., 2014).

Nesse contexto, com a crescente digitalização dos dados em UTIs neonatais, o aprendizado de máquina (AM) surge como uma ferramenta promissora. Algoritmos de AM podem analisar grandes volumes de dados rotineiros do prontuário para identificar padrões e antecipar o reconhecimento da sepsis horas antes da manifestação clínica (KUMAR et al., 2020).

O principal diferencial tecnológico do projeto reside na extração de dados críticos do parto, que não estavam padronizados e se encontravam em formato de texto livre na anamnese médica. Para superar esse desafio, foi implementado um pipeline de Processamento de Linguagem Natural (PLN) utilizando modelos de linguagem de grande escala (LLMs), como Gemma3 e Qwen3, executados localmente através da plataforma Ollama. Essa abordagem permitiu transformar texto não estruturado em variáveis analíticas confiáveis, enriquecendo o conjunto de dados para a predição da sepsis precoce, que é fortemente influenciada pelas condições do parto.

2. ANÁLISE DE MERCADO

A inovação atende a uma necessidade crítica da Unidade de Terapia Intensiva Neonatal (UTIN), nosso público-alvo primário. Nesse ambiente de alta complexidade, as equipes de saúde precisam de ferramentas que ofereçam agilidade diagnóstica e precisão clínica para recém-nascidos críticos. A solução proposta aborda diretamente a limitação dos métodos atuais, que dependem de culturas microbiológicas demoradas, fornecendo um suporte inteligente à decisão que visa reduzir o tempo para intervenção e promover o uso racional de antimicrobianos.

Uma análise do cenário tecnológico e de mercado revela uma clara oportunidade. A pesquisa, que incluiu a base de patentes do INPI, não encontrou soluções de aprendizado de máquina desenvolvidas especificamente para a predição de sepse neonatal. As ferramentas existentes ou em desenvolvimento são focadas na população adulta ou empregam sistemas de pontuação clínica como o MEWS, que não possuem a mesma capacidade preditiva ou especialização para as particularidades do paciente neonato. Essa lacuna posiciona a inovação como uma abordagem pioneira no setor.

A estratégia de entrada no mercado foi desenhada de forma escalonada. A primeira fase consiste na implementação e validação da ferramenta no Hospital Escola da UFPEL, um ambiente de validação controlada dentro da Rede Ebserh. O sucesso neste projeto piloto servirá como um caso de uso robusto para a expansão da tecnologia para outros hospitais da rede. Em uma fase posterior, a solução poderá ser comercializada para hospitais públicos e privados que buscam modernizar seus serviços de neonatologia através da incorporação de inteligência artificial, um mercado com grande potencial de crescimento.

3. ESTRATÉGIA DE DESENVOLVIMENTO E IMPLEMENTAÇÃO

Inicialmente, a inovação é implementada como um projeto de pesquisa e desenvolvimento no Hospital Escola da Universidade Federal de Pelotas, um modelo de inovação social sem expectativa de geração de receita nesta fase. O foco é a validação clínica e o aprimoramento dos algoritmos em um ambiente real. Futuramente, com a validação do protótipo, existe a possibilidade de expansão para hospitais privados, o que poderia levar à criação de um modelo de negócios para comercialização.

Não há previsão para registro de propriedade intelectual. O projeto utiliza modelos de aprendizado de máquina abertos e tem como objetivo a disseminação do conhecimento, permitindo que os códigos e métodos desenvolvidos sejam aproveitados e replicados pela comunidade científica e por outras instituições de saúde.

O desenvolvimento da inovação seguiu etapas bem definidas. A primeira fase consistiu na construção de uma base de dados estruturada a partir de prontuários eletrônicos retrospectivos. Em seguida, foram treinados e comparados cinco algoritmos de aprendizado de máquina (SVM, Random Forest, XGBoost, KNN e MLP) utilizando validação cruzada para identificar o modelo com melhor desempenho. O modelo Random Forest foi selecionado como o protótipo inicial, apresentando a melhor performance (Acurácia 0,75; F1-score 0,692; AUC 0,789). As próximas etapas planejadas incluem a validação prospectiva do modelo com predição diária na UTIN através de um protótipo funcional, além da calibração e análise de curvas de decisão para ajustar os limiares de alerta.

Com base nesse progresso, o nível de maturidade tecnológica da inovação é estimado como **TRL 4**, pois a tecnologia foi validada em um ambiente de testes analíticos, utilizando dados clínicos reais do hospital, demonstrando que a tecnologia possui viabilidade para avançar para a etapa de prova de conceito, com a construção de um protótipo funcional que será testado e implantado no ambiente hospitalar.

O principal desafio técnico foi a extração de informações críticas do parto, que estavam em formato de texto livre. Esse risco foi mitigado com sucesso através do desenvolvimento de um pipeline de PLN com LLMs locais, que estruturou os dados de forma confiável. Outros desafios incluíram o tratamento de dados faltantes e o desbalanceamento de classes (casos de sepse são minoria), que foram gerenciados com técnicas de imputação e o uso de métricas de avaliação robustas, como F1-score e AUC, adequadas para dados desbalanceados.

4. RESULTADOS ESPERADOS E IMPACTO

O principal impacto esperado é social e clínico, com o potencial de reduzir significativamente a morbimortalidade associada à sepse neonatal, especialmente em recém-nascidos prematuros. Ao permitir uma detecção mais rápida e precisa, a inovação contribui para a melhoria da qualidade do atendimento na UTIN. Economicamente, a ferramenta pode diminuir custos hospitalares ao otimizar o uso de antibióticos e, potencialmente, reduzir o tempo de internação.

Na fase atual de pesquisa e validação, a inovação não possui projeções financeiras, pois o foco é o impacto social e a melhoria do processo assistencial no âmbito do SUS. No entanto, a validação bem-sucedida do protótipo abrirá portas para futuras parcerias e modelos de comercialização com instituições privadas, criando um potencial de sustentabilidade financeira a longo prazo.

A visão de futuro para a inovação é a sua evolução para uma ferramenta totalmente integrada ao prontuário eletrônico do hospital, fornecendo alertas de risco em tempo real para a equipe clínica. Com o tempo, o aprimoramento contínuo dos algoritmos permitirá não apenas uma maior precisão na predição da sepse, mas também a expansão do seu uso para o monitoramento de outras condições críticas que exigem diagnóstico ágil. A longo prazo, a solução visa se consolidar como uma tecnologia que moderniza a gestão de dados clínicos, melhora os tratamentos antimicrobianos e contribui para uma gestão hospitalar mais eficiente e sustentável.

5. CONCLUSÕES

A predição de sepse neonatal com aprendizado de máquina representa uma inovação com grande potencial de impacto na saúde neonatal. A validação de um modelo preditivo robusto, como o *Random Forest*, e a aplicação de tecnologias de ponta como LLMs para a estruturação de dados demonstram a viabilidade de se criar uma ferramenta de apoio à decisão clínica eficaz e de baixo custo, com potencial para reduzir a mortalidade e otimizar os recursos do sistema de saúde.

O desenvolvimento e a implementação desta iniciativa dependem de uma rede colaborativa de profissionais e investidores com interesse na temática. O projeto permanece aberto à participação de interessados em contribuir com conhecimentos e recursos para os próximos passos, que incluem a validação prospectiva e a integração da ferramenta ao fluxo assistencial. O objetivo é fomentar uma transformação positiva e duradoura no cuidado neonatal, aliando a expertise clínica à aplicação da inteligência artificial.

6. REFERÊNCIAS BIBLIOGRÁFICAS

KUMAR N, AKANGIRE G, SULLIVAN B, FAIRCHILD K, SAMPATH V. Continuous vital sign analysis for predicting and preventing neonatal diseases in the twenty-first century: big data to the forefront. *Pediatr Res*. 2020; 87: 210-220

MANI S, OZDAS A, ALIFERIS C, VAROL HA, CHEN Q, CARNEVALE R, CHEN Y, ROMANO-KEELER J, NIAN H, WEITKAMP JH. Medical decision support using machine learning for early detection of late-onset neonatal sepsis. *J Am Med Inform Assoc*. 2014 Mar-Apr;21(2):326-36.

MITHAL, L. B.; GERNAND, A. D.; BECK, C.; GALLAGHER, K.; TAYLOR, L. A.; GOLDSTEIN, J. A. Chorioamnionitis and Risk for Maternal and Neonatal Sepsis: A Systematic Review and Meta-analysis. *Obstetrics & Gynecology*, v. 137, n. 6, p. 1007–1022, 2021.

WANG, T.; HANSEN, K.R.; LOVING, J.; PASCHALIDIS, I.C.; AGGELEN, H.V.; SIMHON, E. Predicting Antimicrobial Resistance in the Intensive Care Unit. @inproceedings {Wang2021 PredictingAR, 2021.

WHO, World Health Organization. Global report on infection prevention and control: Executive summary. 2022.