

# Combinando Ontologia e Deep Learning para Integração de Dados de Sensores e Predição de Eventos Críticos

Jefferson Amará<sup>1</sup>, Victor Ströele<sup>1</sup>, Regina Braga<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação (UFJF) – Brasil

jefferson.n.amara@gmail.com, {victor.stroele, regina.braga}@ufjf.br

**Abstract.** *Accurate event detection and prediction are vital for early warning systems to prevent disasters and mitigate risks. This study proposes a framework that combines ontology-based sensor data integration with deep learning to improve critical event forecasting. An extended Semantic Sensor Network (SSN) ontology is used to enrich and standardize heterogeneous sensor data with semantic and contextual information, which then feed a Long Short-Term Memory (LSTM) network for prediction. A case study with real hydrometric and pluviometric data showed that this combined approach improves prediction accuracy and enables earlier detection of critical events compared to isolated sensor analyses.*

**Resumo.** *A detecção e a previsão precisas de eventos são fundamentais para sistemas de alerta precoce, visando prevenir desastres e mitigar riscos. Este estudo propõe um framework que combina integração de dados de sensores baseada em ontologia com aprendizado profundo para aprimorar a previsão de eventos críticos. Uma extensão da ontologia Semantic Sensor Network (SSN) é utilizada para padronizar dados heterogêneos de sensores e enriquecê-los com informações semânticas e contextuais, que então alimentam uma rede Long Short-Term Memory (LSTM) para realizar a previsão. Um estudo de caso com dados reais hidrométricos e pluviométricos demonstrou que essa abordagem combinada melhora a acurácia das previsões e possibilita a detecção antecipada de eventos críticos em comparação com análises isoladas de sensores.*

## 1. Dados da Defesa da Tese e Pontos de Destaque

Data da Defesa: 27 de agosto de 2024

Categoria: Mestrado

Membros da banca:

- Prof. Dr. Victor Ströele de Andrade Menezes (UFJF)
- Prof<sup>a</sup>. Dra. Regina Maria Maciel Braga (UFJF)
- Prof. Dr. José Maria Nazar David (UFJF)
- Prof. Dr. Eduardo Soares Ogasawara (CEFET-RJ)

Orientador: Prof. Dr. Victor Ströele de Andrade Menezes

Co-orientador: Prof<sup>a</sup>. Dra. Regina Maria Maciel Braga

Pontos de Destaque:

- A integração de sensores baseada em ontologia melhora a acurácia das previsões de eventos críticos.
- A ontologia Semantic Sensor Network foi estendida para contextualização de

*dados hidrológicos.*

- *Aprendizado profundo com LSTM aplicado com sucesso para predição antecipada de enchentes.*
- *Framework validado com conjuntos de dados reais de sensores hidrométricos e hidrológicos.*

*Regras contextuais aprimoram a detecção precoce de eventos hidrológicos correlacionados.*

## 2. Contexto e Problema

Sistemas de alerta precoce dependem da detecção e predição antecipada de eventos críticos (como desastres naturais) para prevenir danos. Em domínios como gestão de desastres e cidades inteligentes, há uma abundância de dados de sensores heterogêneos (pluviômetros, sensores de nível de rios, etc.), porém esses dados brutos frequentemente carecem de contexto semântico para uma decisão eficaz. O problema central abordado neste trabalho é como integrar e contextualizar dados de múltiplos sensores heterogêneos de forma a melhorar a identificação de eventos críticos. Trabalhos prévios ressaltam a necessidade de lidar com dependências complexas entre eventos e fontes de dados diversas [Zhao, L. 2021]. No cenário de enchentes, por exemplo, a ocorrência de chuvas intensas em uma região pode influenciar o nível dos rios em outra; sem integração adequada, sinais de alerta podem passar despercebidos ou chegar tarde demais.

## 3. Objetivo

O objetivo deste trabalho é aprimorar a detecção e predição de eventos críticos por meio de um framework que combina integração de dados de sensores baseada em ontologia com técnicas de aprendizado profundo. Em termos específicos, buscou-se permitir que dados heterogêneos (de origens, formatos, grandezas, finalidades e protocolos de comunicação variados) de sensores fossem unificados em um modelo semântico comum e, a partir desse dado enriquecido, aplicar modelos de rede neural recorrente (LSTM) para prever eventos de interesse (como enchentes) com antecedência. Assim, visa-se identificar eventos críticos correlacionados entre diferentes fontes de dados, que poderiam ser ignorados se analisados isoladamente, aumentando a acurácia e a antecedência das previsões.

## 4. Contribuição

As principais contribuições do estudo incluem:

**Integração semântica de sensores:** Desenvolvimento de uma extensão da ontologia Semantic Sensor Network (SSN) [Compton, M. et al. 2012] para incorporar dados hidrológicos e de contexto, permitindo padronizar e enriquecer as leituras de sensores com significado semântico (por exemplo, relacionando medidas de chuva e nível de rio a eventos como enchentes).

**Framework híbrido ontologia-IA:** Proposta de um framework em camadas que combina a ontologia estendida para detecção de eventos via regras de contexto e um modelo de Deep Learning (LSTM) para predição antecipada desses eventos. Essa arquitetura híbrida explora a sinergia entre ontologia e IA: a ontologia detecta eventos críticos com base em limiares e contexto, enquanto o modelo LSTM antecipa tendências futuras usando os dados contextualizados, cada componente cobrindo possíveis falhas do outro.

**Validação em caso real:** Implementação do framework e avaliação em um estudo de caso no domínio hidrológico, integrando dados reais de estações hidrométricas e pluviométricas. Os resultados demonstraram melhoria na acurácia das previsões e capacidade de detecção precoce de enchentes em comparação com abordagens que não utilizam integração semântica, comprovando a efetividade da abordagem proposta.

## 5. Avanço no Estado da Arte

Este trabalho avança o estado da arte ao unir integração semântica de dados de sensores com aprendizado profundo em um só framework para predição de eventos críticos. Estudos

recentes já exploraram predições de enchentes com IA ou integração de sensores, mas com limitações: por exemplo, Al Qundus et al. (2022) utilizaram redes de sensores e SVM para detectar enchentes, porém com limiares específicos pouco generalizáveis [Al Qundus, J. et al. 2022]; Nearing et al. (2024) desenvolveram um modelo global de predição de cheias com LSTMs, mas com menor precisão em contextos locais [Al Qundus, J. et al. 2022]. Em contraste, a solução proposta contextualiza localmente os dados via ontologia e emprega um modelo de IA treinado em dados integrados. Ademais, conforme destacado por Zhao et al. (2021), a consideração de dependências complexas entre eventos é crucial – nossa abordagem ontologia+LSTM captura tais dependências melhor do que técnicas isoladas, permitindo identificar eventos correlatos (por exemplo, enchentes rio abaixo decorrentes de fortes chuvas rio acima) mais cedo e com maior confiabilidade que as abordagens anteriores [Zhao, L. 2021] [Al Qundus, J. et al. 2022] [Nearing, G. et al. 2024].

## 6. Resumo da Solução Proposta

O trabalho propõe um framework em múltiplas camadas para integração de dados de sensores e predição de eventos. Na base, um módulo de aquisição e pré-processamento coleta e limpa dados de diversos sensores (chuva, nível d'água, vazão, etc.). Em seguida, entra a camada de ontologia, onde os dados são mapeados para um esquema conceitual global padronizado – uma ontologia baseada na SSN [Compton, M. et al. 2012] estendida com conceitos hidrológicos. Essa ontologia unifica as diferentes fontes em um modelo comum, adicionando informações de contexto. Regras semânticas e de limiar definidas na ontologia permitem detectar eventos críticos diretamente nos dados integrados (por exemplo, detectar que um dado nível de um rio excede o limiar de enchente em certo ponto do rio). Os dados enriquecidos então alimentam a camada de IA, composta por uma rede LSTM treinada para prever eventos futuros (como a probabilidade de enchente nos próximos dias) com base nas séries temporais. Em resumo, a solução primeiro contextualiza e correlaciona os dados de múltiplos sensores via ontologia e, então, prevê proativamente possíveis eventos críticos via aprendizado profundo, fornecendo um alerta antecipado mais confiável.

## 7. Avaliação

A avaliação consistiu em um estudo de viabilidade usando dados reais de sensores hidrológicos. Foram integrados dados de estações hidrométricas (nível e vazão de rios) e sensores hidrológicos (índice de chuva) de uma bacia hidrográfica, com foco na predição de enchentes. Adotou-se a abordagem GQM (Goal-Question-Metric), definindo como objetivo avaliar se o framework integra semântica e prediz eventos de forma eficaz. Os experimentos mostraram que as métricas de integração semântica foram atendidas – por exemplo, todas as leituras dos diferentes sensores puderam ser mapeadas corretamente para a ontologia (satisfazendo M1), e a ontologia demonstrou flexibilidade para incorporar novos sensores (M2). Na parte de predição, o modelo LSTM treinado com dados integrados obteve alta acurácia na previsão de enchentes (M3), conseguindo antecipar eventos de cheia com maior antecedência comparado a análises isoladas por sensor. Observou-se que a combinação de regras de contexto + LSTM aumenta a sensibilidade: em alguns casos, a ontologia detectou elevações anômalas que serviram de alerta antecipado para o LSTM ajustar sua previsão, enquanto em outros casos o LSTM previu um evento antes dos limiares da ontologia serem alcançados – evidenciando que o duplo mecanismo de detecção/predição reduz falsos negativos. Em suma, a avaliação valida que integrar semanticamente os dados de sensores melhora a qualidade e antecedência das predições de eventos críticos, atendendo ao propósito do estudo.

## 8. Produção Técnica e Científica

- Artigo de evento: “Integrating Heterogeneous Stream and Historical Data Sources using SQL.” *Journal of Information and Data Management (JIDM)* 13(2): 2022. (Este artigo aborda integração de dados heterogêneos via SQL, contribuindo como base para a

presente pesquisa.)

- Artigo em conferência: “Sensor Data Integration Using Ontologies for Event Detection.” In: Proc. 37th IEEE AINA, pp. 171–183, 2023. (Apresenta a abordagem inicial de integração semântica de sensores para detecção de eventos, focando na ontologia SSN estendida.)
- Artigo em conferência: “From Context to Forecast: Ontology-Based Data Integration and AI for Events Prediction.” In: Proc. 38th IEEE AINA, pp. 349–361, 2024. (Descreve o framework completo unindo ontologia e modelo LSTM para predição de enchentes, com resultados experimentais.)
- Framework implementado: Desenvolvimento de um protótipo funcional baseado em microserviços que integra os componentes de ontologia e predição LSTM.
- Dataset anotado: Compilação dos dados brutos de sensores hidrológicos utilizados, enriquecidos com anotações semânticas conforme a ontologia.

## Referências

Compton, M. et al. (2012). The SSN Ontology of the W3C Semantic Sensor Network Incubator Group. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17, 25–32.

Zhao, L. (2021). Event Prediction in the Big Data Era: A Systematic Survey. *ACM Computing Surveys*, 54(5), Article 105.

Al Qundus, J., Dabbour, K., Gupta, S., Meissonier, R., & Paschke, A. (2022). Wireless sensor network for AI-based flood disaster detection. *Proc. of the 55th Hawaii Intl. Conf. on System Sciences (HICSS 2022)*, 4678–4687.

Nearing, G. et al. (2024). Global prediction of extreme floods in ungauged watersheds. *Nature*, 615, 94–100. (Publicado online em 2023, edição impressa de 2024.)