RADemics

Explainable Hybrid AI Models for Risk Assessment in Smart Grid Cybersecurity

Dinesh.P. S, Praveena Rachel Kamala.S, Sangita Gautam Lade.,
BANNARI AMMAN INSTITUTE OF TECHNOLOGY, EASWARI ENGINEERING COLLEGE, VISHWAKARMA INSTITUTE OF TECHNOLOGY

Explainable Hybrid AI Models for Risk Assessment in Smart Grid Cybersecurity

[1]Dinesh.P. S, Assistant Professor Level III, Computer Science and Engineering, Bannari Amman Institute of Technology Sathy - Bhavani State Highway, Alathukombai post, Sathyamangalam, Tamilnadu- 638401, Mail id: dineshpudhu@gmail.com

[2]Praveena Rachel Kamala.S, Assistant Professor, Information technology, Easwari Engineering College, Chennai, Mail ID: praveena.s@eec.srmrmp.edu.in

[3]Sangita Gautam Lade., Assistant Professor, Computer Engineering, Vishwakarma Institute of Technology, Pune, Mail id: sangita.lade@vit.edu

Abstract

The increasing digitization of power infrastructure through smart grid technologies has substantially improved operational efficiency, control, and responsiveness. This transformation has also introduced complex cybersecurity vulnerabilities due to the proliferation of interconnected devices, distributed energy resources (DERs), and real-time control systems. Traditional risk assessment techniques are insufficient to address the dynamic, high-dimensional, and adversarial nature of modern cyber threats. Artificial intelligence (AI) models, particularly those based on machine learning, offer promising solutions for predictive risk detection but often lack interpretability, thereby limiting trust, compliance, and operational usability in critical infrastructure contexts.This chapter presents a comprehensive framework for the design and deployment of explainable hybrid AI models tailored for real-time risk assessment in smart grid cybersecurity environments. The proposed approach integrates symbolic reasoning, data-driven learning, and explainable AI (XAI) methodologies within a modular architecture that enables both high detection accuracy and interpretable decision-making. Emphasis is placed on balancing model complexity with operational transparency, ensuring the system's adaptability across heterogeneous and resource-constrained grid environments. Furthermore, challenges related to explanation latency, scalability, and human interpretability are addressed through lightweight, context-aware XAI techniques embedded within the hybrid inference pipeline.

Keywords: Smart Grid Cybersecurity, Hybrid Artificial Intelligence, Risk Assessment, Explainable AI, Real-Time Detection, Distributed Energy Resources

Introduction

The transition from conventional power systems to intelligent smart grids has redefined the landscape of energy generation, transmission, and distribution [1]. Characterized by the integration of cyber-physical systems, real-time monitoring, and bi-directional data flows, smart grids offer improved energy efficiency, reliability, and sustainability [2]. This rapid digital transformation has significantly expanded the grid's attack surface, introducing critical vulnerabilities across interconnected components such as intelligent substations, advanced metering infrastructure (AMI), distributed energy resources (DERs), and supervisory control and data acquisition

(SCADA) systems [3]. These components are increasingly targeted by cyber adversaries employing sophisticated methods including zero-day exploits, coordinated denial-of-service attacks, and data manipulation strategies [4]. As these systems evolve in complexity, ensuring their cybersecurity becomes imperative for maintaining grid stability, operational continuity, and public safety [5].

The traditional methods used for risk assessment in power grids are often based on static rules, signature-based detection systems, and qualitative expert evaluations [6]. While effective in predictable threat scenarios, these approaches fall short in identifying emerging, polymorphic, or previously unknown attack vectors [7]. Their inability to process large-scale, multi-dimensional data in real time limits their relevance in today's smart grid ecosystems [8]. Artificial intelligence (AI), and more specifically, machine learning (ML) and deep learning (DL), have demonstrated the potential to overcome these limitations. By learning from historical data and adapting to evolving patterns, AI models can detect anomalies, forecast threats, and assess risk with higher precision [9]. The inherent opacity of many advanced AI algorithms presents a major obstacle in critical infrastructure applications, where interpretability, trust, and human oversight are non-negotiable requirements [10].

The lack of transparency in AI-based cybersecurity systems poses significant challenges for stakeholders in smart grid environments [11]. Grid operators, regulatory bodies, and cybersecurity analysts require not only accurate predictions but also clear justifications for the decisions made by automated systems [12]. The black-box nature of deep learning models makes it difficult to trace the reasoning behind specific outcomes, thereby reducing their acceptance in mission-critical scenarios [13]. This issue becomes even more pressing when considering compliance with industry regulations such as the NERC-CIP framework, which mandates auditability and explainability in risk management processes [14]. In the absence of interpretable outputs, the utility of AI in decision-making remains limited, regardless of its statistical accuracy. This disconnect between technical performance and operational usability underscores the need for models that not only perform well but also communicate their logic effectively to human users in real time [15].