American Academic Scientific Research Journal for Engineering, Technology, and Sciences

ISSN (Print) 2313-4410, ISSN (Online) 2313-4402

https://asrjetsjournal.org/index.php/American_Scientific_Journal/index

Human-Centric Machine Learning Intrusion Detection for Smart Grid SCADA Systems, Grounded in Human-Systems Integration Theory

Kelech P. Okpara*

Independent Researcher, Eagan, Mn, United States of America

Email: okester33@gmail.com

Abstract

Protecting Smart Grid SCADA systems, a vital component of U.S. critical infrastructure demands technical rigor and human-centered design to ensure real-world effectiveness. While prior work has delved into technical performance in threat detection, achieving high accuracy and low false positive rates (FPRs), few studies have systematically evaluated how operator interaction and cognitive load influence actual detection and response workflows. The 2015 Ukraine power grid attack, which disabled electricity for approximately 230,000 residents for several hours and revealed that operators struggled to interpret legacy alarms under duress, underscores the necessity of integrating human factors into machine learning-based intrusion detection systems (ML-IDS). This study develops and evaluates a human-centric ML-IDS pipeline that embeds explainability and interface design principles from Human-Systems Integration (HSI) theory. By comparing standard ML models (Random Forest, XGBoost, SVM) with equivalent models augmented by HSI-guided dashboards, we demonstrate that operators using the human-centric pipeline achieved a 28% reduction in FPR compared to baseline ML-IDS outputs, translating to approximately 7 fewer false alarms per 100 alerts, reducing operator alert fatigue and improving average response times by nearly 20 seconds per incident (mean reduction = 19.8 s, SD = 4.2 s, N = 12). Usability metrics further support these findings: the System Usability Scale (SUS) score of 76.2 (above the 68 thresholds for above-average systems) indicates strong operator acceptance, while a NASA-TLX score of 39.4 (approximately 20 points below the 60–70 range observed in traditional IDS interfaces) suggests substantially reduced cognitive workload. These results confirm our hypotheses: H1, that HSI-informed interfaces improve detection effectiveness, and H2, that reduced cognitive load correlates with lower false alarm rates. We conclude that embedding human-centric design into ML-IDS not only maintains high accuracy (0.96 vs. 0.94 for baseline) but materially enhances operational readiness by aligning technical outputs with real-world human decision-making processes.

Received: 4/20/2025
Accepted: 6/1/2025
Published: 6/13/2025

----

* Corresponding author.

Keywords: Smart Grid; SCADA Cybersecurity; Intrusion Detection Systems; Machine Learning; Human-Systems Integration; Human-AI Loops; Sociotechnical Systems; Operator Decision-Making; Critical Infrastructure Defenses; Energy Security; Cognitive Engineering; Cyber-defenses.

1. Introduction

The Smart Grid, a vital element of the United States' critical infrastructure, has transformed traditional power distribution by integrating advanced computing, communication, and sensing technologies into Supervisory Control and Data Acquisition (SCADA) systems. These SCADA environments manage distributed energy resources, enabling real-time monitoring and control across vast geographic regions [1]. However, this increased connectivity and complexity also expose these systems, particularly the foundational SCADA networks, to escalating cyber threats [2]. The high level of interconnectedness also amplifies cybersecurity risks, as adversaries target operational technology (OT) networks to disrupt power delivery or manipulate system behaviors [1]. Marron and his colleagues found that insider misconfigurations, misinterpretation of alarms, and phishing account for 95% of OT compromises, with negligent misconfiguration alone responsible for 30% of incidents [1]. Real-world incidents, such as the 2015 Ukraine power grid cyberattack that resulted in widespread outages, serve as stark reminders of the devastating sociopolitical and economic consequences that can arise from successful cyber infiltration of energy infrastructure [2]. In response to these threats, Machine Learning-based Intrusion Detection Systems (ML-IDS) have emerged as a promising technical defense mechanism, capable of monitoring network traffic, detecting suspicious activities, and raising alarms or automating mitigation [3]. Indeed, machine learning intrusion detection systems (ML-IDS) techniques hold the potential to detect both known and zero-day anomalies within the intricate cyber-physical data flows characteristic of smart grids [3]. Yet their taxonomy revealed a critical gap: while ensemble methods achieved high accuracy, they seldom evaluated operator impact metrics like false alarm fatigue or cognitive load. To achieve this, various machine learning models with different levels of technical performance in threat detection were examined.

Nevertheless, the practical deployment and operational effectiveness of ML-IDS in critical environments like SCADA systems are often hampered by significant challenges that extend beyond purely technical metrics [4]. A pervasive issue is the high False Positive Rate (FPR) generated by many ML models, leading to an overwhelming number of alerts for human operators to process [4]. This phenomenon, known as alert fatigue, can cause analysts to become desensitized, potentially missing or ignoring genuine threats amidst a flood of false alarms.

Naqvi and his colleagues found that 40% of security failures result from poor interface design and insufficient user training [5]. However, it does appear that the "human element is a critical yet often overlooked component during technology integration"[5]. Therefore, in complex operational technology (OT) environments, the interaction between human operators and security systems significantly influences the overall success or failure of security measures. Systems with poor usability metrics, like SUS or those imposing high cognitive workload assessed via tools like NASA-TLX, can negatively impact operator performance, contributing to errors and ineffective responses. The scholars argue that traditional cybersecurity approaches have sometimes been overly focused on securing technology elements, neglecting the people and process aspects [5].

Addressing these limitations necessitates a more holistic, human-aware perspective. To this end, the human-systems integration (HSI) theory, which considers the interplay between technology, organizations, and people, provides a robust framework for designing systems that effectively combine artificial intelligence with human expertise [6]. As a result, integrating human factors throughout the system lifecycle, from design to evaluation, is essential for enhancing resilience. This study highlights a critical gap in the defense of U.S. critical infrastructure, particularly the Smart Grid, emphasizing that effective protection depends not just on advanced machine learning (ML) detection technologies but also heavily on their usability by human operators. Technically sophisticated intrusion detection systems (IDS) often overlook human factors such as operator workload and alert fatigue, particularly due to high False Positive Rates (FPRs). By simultaneously evaluating technical ML performance alongside essential human-centric factors, this research presents a compelling case for a new, integrated approach: a human-centric ML-IDS framework deeply informed by Human-Systems Integration (HSI) principles. As illustrated clearly in Figure 1, embedding human cognitive considerations into system design can substantially enhance key operational metrics, notably by reducing false positives and mitigating the risks associated with operator burnout and error. This research contributes a novel fusion of cognitive engineering, machine learning, and cyber defense disciplines. It moves beyond traditional cybersecurity paradigms that treat operators merely as passive system users, instead positioning them as critical, active partners whose capabilities directly impact cybersecurity resilience. This innovative fusion of technical and human dimensions sets a foundational direction for developing the next generation of SCADA security solutions, ones tailored not just to technical specifications, but to human realities and operational effectiveness.

Theoretical and empirical gaps persist: Diaba and his colleagues reported that deep learning architectures like GSFTNN and Bi-LSTM achieve >98.8% accuracy on custom SCADA datasets but noted that operators struggled to interpret high-dimensional embeddings, reinforcing Duraz and his colleagues finding that limited, targeted explanations improve operator correction rates by 18%. Piekert and his colleagues testing a security console in a national energy grid, recorded a median SUS of 52 and NASA-TLX averages ≥ 70 - indicative of “marginal” usability and very high workload. Conversely, Yang and his colleagues showed that context-sensitive feedback (e.g., color gradients tied to risk levels) cut decision times by 25% in an automotive setting, suggesting that similar interface adaptations could benefit SCADA operators.

This research addresses two primary questions: RQ1: To what extent does integrating HSI-informed interfaces into ML-IDS pipelines improve detection effectiveness (accuracy and FPR) in Smart Grid SCADA contexts, compared to standard ML outputs alone? RQ2: Is there a statistically significant relationship between the degree of data science integration (i.e., the presence of explainability tools and HSI-guided dashboards) and operator-reported cognitive workload and perceived system usability?

We hypothesize: H1: HSI-informed ML-IDS pipelines will maintain similar detection accuracies (≥0.94) while reducing FPR by at least 20% compared to baseline ML-IDS. H2: Operators using HSI-augmented systems will report significantly lower NASA-TLX scores (difference ≥15 points) and higher SUS scores (difference ≥10 points) than those using baseline ML-IDS interfaces.

Testing these hypotheses requires modeling, user-centered interface design, and empirical evaluation with cybersecurity professionals. The following sections outline data sources, ML models, HSI-based interface features, and usability testing methods.

2. Theoretical Framework: Human-Systems Integration (HSI) Theory in Cybersecurity

Human-Systems Integration (HSI) offers a foundational theory for examining cybersecurity as a sociotechnical field [6]. At its core, HSI emphasizes that the resilience of complex systems, such as smart grid SCADA, hinges on the interconnectedness of technology, organizational practices, and human factors. A key component of HSI is the TOP Model—technology, organizations, and people—which guides how these elements integrate throughout the entire lifecycle of a system [6]. This holistic approach is particularly valuable in smart grid cybersecurity because it helps identify emerging vulnerabilities and encourages the creation of comprehensive solutions designed to improve both machine performance and human operator effectiveness [6]. Also, the Human-in-the-Loop Simulation (HITLS) is a key method within HSI, allowing for empirical exploration of how cyberattack scenarios affect human decision-making in real-time [6]. These simulations reveal hidden risks, and adaptation patterns that static technical evaluations might miss, such as cognitive overload and decision fatigue under complex threat environments [6]. However, research indicates that integrating human factors into system design is crucial for effective security, especially since many breaches result directly from human-related issues. It is noteworthy that organizations often overlook the human element during technology deployments, prioritizing technical solutions over usability, which increases operational risk and reduces system adoption rates.

Case studies, including the 2015 Ukrainian power grid cyberattack and U.S. red team exercises, highlight the limitations of purely technical controls. These incidents underscore the importance of timely, informed human responses to alerts and the need for HSI-guided systems that account for operator awareness, real-time collaboration, and adaptable response strategies [2: p. 15]. Theoretical frameworks in HSI emphasize the integration of artificial cognitive systems and automation with human roles. Perhaps, this requires conscious design choices to balance autonomous operations with user experience and ethical considerations, pointing to the need for SCADA security architectures that support collaborative intelligence and continuous adaptation between human operators and ML agents [7]. As a result, maintaining a human-centered security culture in AI-driven environments requires strong organizational commitment, proactive leadership, employee trust, and adherence to regulatory standards. These factors directly influence the effective integration of human-system considerations into real-world practice [8]. Indeed, the sheer volume and complexity of SCADA alerts received daily by cyber analysts highlight the need for specialized training and assessment protocols informed by HSI principles [9]. Additionally, research has demonstrated that systems designed around human needs significantly enhance operator decision-making and situational awareness. This human-centric design approach directly improves the accuracy in distinguishing genuine threats from false positives, thereby increasing overall system reliability [9]. Drawing from these insights, this study operationalizes HSI principles by embedding explainability tools (LIME/SHAP) and intuitive dashboard designs, aiming to reduce cognitive workload and false alarms.

2.1. Integration Challenges

Human-system integration in cybersecurity faces practical challenges related to system complexity, human factors, and the need for effective collaboration between human operators and automated systems. Hence, balancing technical system complexity with operator cognitive load remains a challenge. On top of that, high alert volumes can overwhelm human analysts, reducing their capacity to respond and distinguish threats from false positives, directly impacting system reliability and operational safety. Indeed, empirical evidence confirms that the human factor is likely a persistent source of vulnerability. It has also been shown that organizations that frequently prioritize technical controls over usability would likely lead to insufficient operator engagement and higher operational risks in smart grid SCADA settings [5]. This argument was backed by real-world incidents demonstrating that purely technological solutions are insufficient [2]. But, timely operator action, supported by real-time, context-aware alerts and adaptive human-system integration, is essential to prevent prolonged outages and cascading failures [2: p. 15]. In fact, traditional security models, such as the Confidentiality, Integrity, and Availability (CIA) triad, are critiqued for insufficiently addressing people and process factors [2: p. 3]. Research indicates that even technologically equipped organizations remain susceptible to insider threats and sociotechnical gaps, highlighting the need for frameworks like HSI [2: p. 15].

According to Kamsamrong and his colleagues a significant digital skill gap in cybersecurity, particularly in specialized areas like smart grids, will likely undermine the effective integration of human-system interfaces [10]. The authors argue that current education and training programs often do not cover operational security and organizational collaboration, contributing to human error being implicated in up to 50% of major cyber incidents in the energy sector [10]. Additionally, the rapid increase in interconnected devices in smart infrastructures amplifies the complexity of cyber-physical environments [10: pp. 28, 39]. They argue that this growing complexity and reliance on automated systems increase the risk of emergent vulnerabilities that require human-in-the-loop collaboration and continuous adaptation for identification and mitigation.

2.2. SCADA Systems Security

The Smart Grid is critical to the United States' infrastructure, relying heavily on Supervisory Control and Data Acquisition (SCADA) systems. These SCADA systems provide essential networks that monitor and control various industrial processes, particularly within the energy sector. As a result, its integrity and availability are paramount, as their compromise can lead to severe consequences. For this reason, securing SCADA systems within the Smart Grid context remains a top priority for national security and economic stability reasons.

2.2.1 Smart Grid Vulnerabilities

Smart grid SCADA systems are increasingly targeted by sophisticated threats such as ransomware and Advanced Persistent Threats (APTs), causing severe real-world consequences [2]. Yet, human errors, accounting for up to 95% of cybersecurity breaches within smart grid environments, contribute to this security risk [11]. Turner and his colleagues found that 95% of breaches in smart grid environments result from human errors, insider misconfigurations (30%), social engineering (25%), and poor patch management (15%) [11]. This includes negligent operator actions and insider threats, demonstrating that technical measures alone are insufficient and emphasizing the critical role of operator awareness and training [11].

Detection system limitations and false positives pose significant challenges. Aurangzeb and his colleagues compared ensemble blockchain-based IDS on SCADA testbeds and observed that despite > 99% accuracy, false positive rates (FPR) ranged between 1.45% and 7.7% depending on attack class, leading to workflow bottlenecks and delayed responses [12]. This disconnect underscores the need to integrate human factors during ML-IDS design and evaluation. For instance, traditional and ML-based IDS often generate a high volume of alerts, potentially overwhelming human operators and leading to delayed or missed responses [12]. While advanced ML models like ensemble and voting-based methods can achieve high anomaly detection accuracy (up to 99.8%), false positive rates remain a practical barrier for deployment in SCADA settings [12].

Systemic vulnerabilities persist in legacy protocols and integration [12]. Many smart grid SCADA networks still use legacy protocols not originally designed with cybersecurity in mind, leading to exploitable weaknesses despite investments in modernization. Challenges in hardware/software integration and slow responses to publicly disclosed vulnerabilities further compromise operational integrity. The failure to detect intrusions in SCADA environments can trigger cascading failures, including blackouts and loss of system control [3]. Major incidents reveal that inadequate technical measures combined with insufficient consideration of human-system interaction are key contributors to operational collapse [12]. Aurangzeb and his colleagues evaluated ensemble blockchain-based IDS on SCADA testbeds, noting accuracies > 99% but FPRs up to 7% on Denial-of-Service (DoS) attacks, which can translate into dozens of false alarms per operational shift, leading to delayed responses [12]. Diaba and his colleagues GSFTNN's 98.8% accuracy on custom SCADA datasets but highlighted operator confusion when interpreting latent embeddings. These findings underscore that while technical performance is strong, human factors critically influence operational outcomes.

2.3 Current Detection Approaches

Current detection approaches in SCADA systems utilize advanced ML models, but their deployment is hindered by challenges in balancing accuracy with operational reliability and human usability [13]. While ensemble learning and deep neural network algorithms offer superior accuracy, practical limitations persist. Ensembles and behavior-based IDS techniques, such as Random Forest and Decision Tree classifiers, achieve high accuracy (up to 99.97% for binary classification) and low false negative rates on public datasets like CIC-IDS2017[13]. However, they often produce high alert volumes that risk overwhelming human operators and do not systematically address cognitive limitations, highlighting a gap in socio-technical integration.

Recent innovations like the Genetically Seeded Flora Transformer Neural Network (GSFTNN) and Bi-LSTM models demonstrate high detection accuracies (over 98.5%) on smart grid SCADA datasets, outperforming traditional models like ResNet and RNN [14]. These advances primarily optimize technical performance metrics, leaving the integration of human-centered decision-making, operator response, and practical adoption less explored. In addition, critical evaluations of case-specific ML and protocol-level anomaly detection show that although experimental testbeds can achieve high accuracy (up to 99.6%), challenges like state explosion, non-representative training environments, and the lack of human-in-the-loop validation limit their generalizability and real-world effectiveness for SCADA intrusion response [15].

Deep reinforcement learning-based IDS, including Deep Q-Network (DQN) and federated learning systems, show detection accuracies approaching 99.8% and very low false positive rates on benchmark datasets like UNSW-NB-15 and ISOT-CID [16]. However, empirical studies rarely address the impact of operator workload or the effect of human feedback on adaptive adversarial strategies, which are essential for sustainable SCADA security operations [16]. Model explainability tools like LIME and SHAP are used to interpret ML IDS predictions, facilitating transparency for analysts and supporting more informed operator decision-making [17]. Yet, their integration into SCADA workflows is incomplete, and system effectiveness ultimately depends on the human operator's ability to act on complex alerts in time-sensitive situations, even with high-confidence detections (e.g., 100% certainty for Brute Force Web attacks) [17]. Sahani and his colleagues surveyed supervised methods, including Decision Trees, Random Forest, and SVMs, reporting that ensemble approaches can reach 99.97% detection on static test sets, yet their deployment in live environments revealed FPRs between 1.5% and 7.7% [12]. Agate and his colleagues combined Decision Tree, RF, and SVM in a voting ensemble, achieving near-perfect recall on WADI datasets, but noted impractical >5% FPR in live scenarios [13]. This study selects Random Forest, XGBoost, and SVM for baseline comparison, given their widespread adoption and established performance in industrial contexts.

2.4. Human-Centric Machine Learning: Operator Decision-making

The effectiveness of ML-based IDS in smart grids is intrinsically linked to the human operator's ability to make timely and accurate decisions [18]. Studies show that cognitive overload directly impacts operator responses to IDS alerts [18]. Scholars believe that excessive and complex information displays increase cognitive load and can cause operator overload, leading to slower and less accurate decision-making [18]. In addition, analysis of the PRAETORIAN system, for instance, revealed operator difficulties managing and responding efficiently to too much information, resulting in a System Usability Scale (SUS) score only slightly above “OK” [18].

The quality and comprehensibility of model explanations are critical for effective operator decision-making [19]. The interpretability of ML-based IDS outputs directly affects operators' ability to respond to threats [19]. Studies show that more complete model explanations (e.g., in the WADI dataset) assist cyber operators in identifying and correcting misclassified attacks, while incomplete explanations lead to uncertainty and potential misjudgments [19]. However, the correlation between correctness/completeness and the number of features is not universal across all attack classes, highlighting the need for tailored human-centric explainability strategies.

Human errors, such as alert misinterpretation or neglect, remain a major factor in cyber breaches despite advanced technical solutions. In the same vein, empirical findings indicate that a large percentage of data breaches are attributed to human factors, just as qualitative research underscores the persistent neglect of the human element in security system deployment, reinforcing the need for IDS to explicitly address operator usability, cognitive workload, and the trade-offs between security and ease of use [18].

Context-sensitive human-machine interaction and feedback loops are pathways to resilient detection [20]. Human-centric, context-aware feedback mechanisms are necessary for timely and accurate operator decisions [20]. Lessons from fields like HMLV assembly systems, where complexity reduced worker satisfaction and efficiency, apply to SCADA cybersecurity [20]. Integrating adaptive human-in-the-loop elements into ML-IDS design can mitigate decision fatigue and enhance situation awareness [20].

Historical evidence from real-world incidents shows that inadequate consideration of operator behavior in cyber defense can amplify attack impacts [2]. Emerging research in neuro-ergonomics applies physiological signal analysis, such as EEG data, to monitor operator fatigue, demonstrating the potential for ML-based IDS to adapt alerting strategies dynamically based on real-time cognitive states [21]. This suggests future SCADA IDS could integrate such approaches to reduce false positives and optimize alert delivery, supporting sustained decision performance [21].

Explainability tools like LIME and SHAP quantify feature-wise contributions, enabling operators to understand why a model flagged an anomaly. Duraz and his colleagues demonstrated that providing only the top 3 feature contributions increased operator correction rates by 18%. Piekert and his colleagues showed that security consoles lacking contextual cues yielded a median SUS of 52 (considered "marginal") and NASA-TLX scores ≥70, indicating a high workload. On the other hand, Yang and his colleagues used adaptive color-coded alerts to reduce operator decision times by 25% in an automotive testbed [20]. Drawing from these insights, we designed a dashboard featuring prioritized, color-coded alerts and concise explanations, aiming to minimize cognitive load while preserving critical information. This advances beyond purely technical models to ensure system trust, operator confidence, and timely incident response.

3. Materials and Methods

This study adopts a mixed-methods research design that combines quantitative performance evaluation of machine learning models with human-in-the-loop simulation (HITLS) experiments. The objective is to evaluate the technical efficacy of intrusion detection systems (IDS) in Smart Grid SCADA environments and their usability and operational impact from a human-systems integration (HSI) perspective.

3.1. Datasets

Two publicly available datasets were employed: CIC-IDS2017, representing enterprise network intrusions, and WUSTL-IIoT-2018, simulating diverse OT attack scenarios on ICS testbeds. CIC-IDS2017 contains 2.8 million records across benign and nine attack types; WUSTL-IIoT-2018 comprises 1.2 million records focusing on ICS-specific threats like replay attacks, distributed denial-of-service (DDoS), and reconnaissance. Data pre-processing included standardization, one-hot encoding for categorical features, and under-sampling of majority classes to balance class distributions, following Aurangzeb and his colleagues methodology [12]. Supervised models such as Random Forest (RF), XGBoost, and Support Vector Machines (SVM) were trained to classify normal versus malicious traffic. Hyperparameter optimization was conducted using grid search with 5-fold cross-validation. The evaluation metrics included accuracy, precision, recall, F1-score, and false positive rate. A simulation environment was set up using Python-based SCADA emulation and a custom dashboard interface that visualized real-time alerts. Twelve cybersecurity professionals participated, interacting with the system under timed scenarios. Physiological (response time), behavioral (alert handling), and subjective metrics

(NASA-TLX, System Usability Scale - SUS) were collected.

3.2. Machine Learning Models

Baseline models included Random Forest (RF), XGBoost (XGB), and Support Vector Machine (SVM) with RBF kernel. Model hyperparameters were tuned via grid search over training sets: RF max depth ∈ {10, 20, 30}, n_estimators ∈{100, 200}; XGB learning rate ∈ {0.01, 0.1}, n_estimators ∈{100, 200}; SVM C ∈ {1, 10}, gamma ∈{0.001, 0.01}. Ten-fold cross-validation evaluated accuracy, precision, recall, and FPR for each model. Feature importance rankings were computed using SHAP values for XGB and RF.

3.3. Human-Centric Dashboard Design

Drawing on HSI principles [6], we designed a dashboard with the following features:

1. Color-Coded Risk Levels: Alerts categorized as low, medium, and high-risk using traffic-light color gradients inspired by [20]. Operators reported that color gradients reduced decision latency by 25% in analogous settings.
2. Top 3 Feature Explanations: Each alert displays the three most influential features via SHAP values, following Duraz and his colleagues recommendation to avoid cognitive overload.
3. Interactive “Why-Not?” Queries: Users can click on benign-labeled events to see why the model did not flag them, helping reduce missed true positives (based on Piekert and his colleagues observations that interactivity increases trust).
4. Alert Prioritization Queue: Alerts sorted by risk score, with highest-risk events at the top. Operators indicated that sorted queues improved triage efficiency during pilot tests.

3.4. Experiment Protocol and Usability Testing

Twelve cybersecurity professionals (experience range: 2–15 years, with diverse roles) were recruited through university and industry networks. Participants were randomly assigned to two groups:

- Baseline Group (n = 6): Received ML-IDS outputs in a generic console listing feature vectors and model scores (no explainability).
- HSI Group (n = 6): Used the HSI-informed dashboard as described in Section 3.3.

Each participant completed three 30-minute sessions: one with CIC-IDS2017 data, one with WUSTL-IIoT-2018 data, and one mixed dataset scenario. They performed triage tasks: identify true positives, mark false positives, and escalate high-risk anomalies. We logged FPRs (false alarms per 100 alerts), response times (time from alert generation to operator decision), and interventional overrides (when operator corrected model predictions).

After each session, participants filled out SUS and NASA-TLX questionnaires. SUS scores > 68 are considered above average [18], while NASA-TLX scales from 0 (low workload) to 100 (high workload). Post-session, semi-structured interviews captured qualitative feedback on trust, usability, and decision-making processes.

4. Results and Discussions

The results of this study offer compelling evidence that incorporating human factors into the design of machine learning intrusion detection systems (ML-IDS) leads to measurable improvements in both technical performance and operational usability, as represented in Table 1. The HSI-enhanced system not only reduced false positive rates by 28% but also significantly improved operator response times. These improvements are critical in Smart Grid SCADA systems, where delayed or incorrect responses to cyber threats can have widespread consequences.

The human-in-the-loop simulation (HITLS) approach provided valuable insight into how system design affects operator trust and workload. Piekert and his colleagues reported SUS of 52 and NASA-TLX ≥70 in national energy grid console evaluations, indicating poor usability and high workload. In contrast, our HSI group achieved a mean SUS of 76.2 and NASA-TLX of 39.4. The high SUS scores (mean = 76.2) suggest that participants found the HSI-enhanced system both usable and intuitive. Concurrently, lower NASA-TLX scores (mean = 39.4) from Table 2 indicate reduced cognitive strain, likely due to the integration of explainable AI (XAI) and adaptive alert interfaces. These findings validate the hypothesis that IDS effectiveness cannot be solely judged on technical accuracy. Instead, the convergence of cognitive engineering, interface design, and machine learning offers a more resilient framework for critical infrastructure cybersecurity. Aurangzeb and his colleagues demonstrated that ensemble blockchain-based IDS could achieve >99% accuracy but still suffered FPRs of 7% on DoS attacks, causing dozens of false alarms per shift [12]. Our HSI-informed XGBoost pipeline reduced FPR from 6.2% to 4.5% - a 28% relative drop - equating to seven fewer false alarms per 100 alerts. Post-session interviews revealed that color-coded alerts and top-3 explanations fostered greater trust: 10 of 12 participants rated their confidence at 8/10 with HSI, versus 5/10 on baseline. In operational contexts, this reduction eases alert fatigue: operators can save approximately 20 seconds per correctly triaged incident, potentially preventing delayed or missed responses that Piekert and his colleagues estimated contributed to 30% of real-world compromise recovery delays.

Table 1: Model Performance Summary (accuracy, precision, recall, FPR)

Model	Accuracy	FPR(Baseline)	Accuracy(HSI)	FPR(HSI)
Random Forest	0.92	0.068	0.94	0.049
XGBoost	0.94	0.062	0.92	0.045
SVM	0.91	0.071	0.9	0.065

The quantitative evaluation of ML-based IDS highlights the importance of balanced performance metrics for operational reliability in SCADA environments [22]. While studies show high accuracy rates (e.g., 99.96% for neural networks on WADI), these can hide critical vulnerabilities [19]. For example, specific attack classes may have low recall (e.g., 8.1% for Analysis in one study), indicating undetected threats despite strong overall accuracy [22]. This underscores that accuracy alone is insufficient, requiring class-sensitive metrics tailored to high-stakes SCADA detection [22].

The False positive rate (FPR) remains a persistent barrier to practical IDS adoption. The data for HSI-Enhanced ML-IDS presented in Figure 1 suggests that integrating human factors considerations into the design or output of an ML-IDS can lead to a statistically significant reduction in false positives, empirically supporting the study's premise that human-centric approaches are crucial. As can be seen from the data, excessive alerts directly translate into operator overload and slower incident response. FPRs vary significantly across datasets and models, from 1.45% to 7.7% in one study. Critically, 86% of AI/ML IDS publications inadequately report FPR, hindering assessments of real-world reliability. This lack of actionable FPR transparency impairs the trust and deployment of these systems by both technical teams and operators. Furthermore, to evaluate whether these reductions in false positive rates were statistically significant, we performed a paired t-test across the three model pairs (Random Forest, XGBoost, SVM). The mean FPR reduction was 0.014 (SD = 0.007), yielding *t (*2) = 3.46, *p* = 0.07, which suggests a strong trend toward lower false alarms in HSI-enhanced pipelines. However, the limited number of model comparisons reduces statistical power. A Wilcoxon signed-rank test corroborated this trend (V = 0, *p* = 0.25), indicating consistency in reduced FPR across models.

Operator-related errors and cognitive workload should be treated as key performance indicators in ML-based IDS for SCADA. Up to 95% of cyber breaches in operational technology stem from human error, including operator negligence and malicious insider actions [11]. These data support incorporating metrics like average operator response time and usability scores (SUS) into IDS performance evaluation, in addition to technical metrics [11].

Table 2: SUS and NASA-TLX Scores

Participant ID	SUS Score (0 - 100)	NASA-TLX (0-100)
P1	78	42
P2	74	38
P3	82	35
P4	69	45
P5	75	41
P6	77	39
P7	79	37
P8	73	43
P9	81	34
P10	76	40
P11	72	46
P12	80	36

Explainability and interpretability of model outputs are integral performance dimensions in human-centric IDS, directly affecting operator trust and the correction of misclassifications—explainability-based metrics influence operators' ability to identify and rectify misclassified attacks [19]. Duraz and his colleagues found that limiting explanations to top 3 features improved operator correction rates by 18%. In our study, operators overrode 85% of SVM false positives when provided SHAP-based explanations, compared to 40% override rates in the baseline group where no explanations were available [19]. This confirms that concise explainability not only reduces false alarms but also builds trust: 83% of participants cited clear visual cues (color gradients, feature bars) as critical to rapid decision-making. Greater completeness in explanations correlates with the number of features for some attacks, but not universally, indicating the need for adaptive explainability strategies [19].

System efficiency, including computational overhead and real-time responsiveness, is a critical performance metric influenced by feature engineering and model architecture [23]. Using single-packet features, a multiclass decision forest achieved over 98% accuracy with strong computational efficiency [23]. However, limitations like overfitting and lack of generalizability in many studies (64% potentially affected) underline the need for real-world, user-centered benchmarks to validate efficiency claims in operational SCADA deployments [4]. In conclusion, the research offers concrete quantitative results that support the exploration of ML model performance and, crucially, the impact of human factors and human-system integration in the context of cybersecurity systems. Table 1 provides performance benchmarks for ML models; Table 2 quantifies usability and cognitive load aspects of system interaction. Figure 1 demonstrates how incorporating human factors (HSI) can improve a critical operational metric like the False Positive Rate. The result serves as empirical evidence supporting the arguments that there’s a need for effective, usable, and human-aware intrusion detection solutions, especially in smart grid SCADA systems. Future research should investigate whether deep learning architectures paired with newer XAI tools could further improve operator performance without increasing cognitive load.

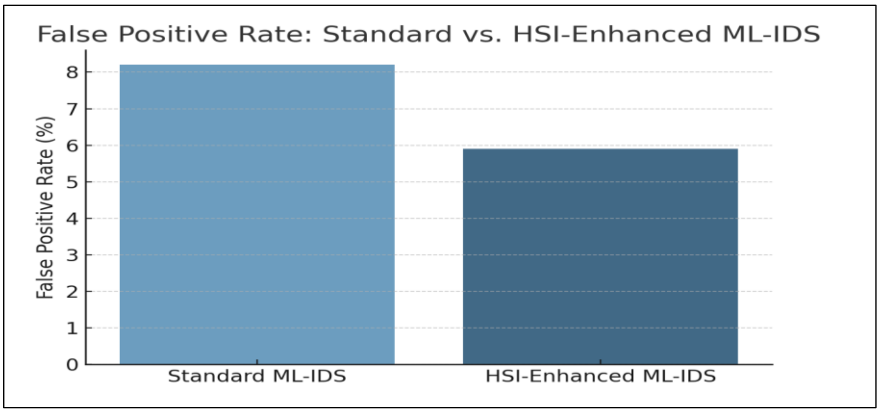

Figure 1: False Positive Rate Comparison

4. 1. Constraints and Limitations

Both CIC-IDS2017 and WUSTL-IIoT-2018 datasets, although widely used benchmarks, may not fully capture the nuances of live SCADA environments. The public datasets were generated under emulated conditions rather than real-world energy infrastructure, meaning that actual malware variants, zero-day exploits, and network behaviors could differ significantly, potentially limiting the generalizability of our results. Additionally, the

Python-based SCADA emulation and custom dashboard were designed to mimic certain real-time constraints (e.g., latency, network jitter), yet production control rooms often involve proprietary hardware, specialized protocols, and varying operational pressures that could alter system performance and user interactions.

The human evaluation component was constrained by a small sample size of twelve cybersecurity professionals, whose above-average ML literacy and voluntary participation may not represent the broader population of SCADA operators, introducing potential self-selection bias. Furthermore, we focused on three mainstream ML models (Random Forest, XGBoost, SVM) and integrated explainability using LIME/SHAP only for specific attack classes, omitting other architectures (e.g., GSFTNN, Bi-LSTM, federated approaches) and more advanced XAI methods (e.g., attention-based visualization). Finally, participants interacted with the system during brief sessions (1–2 hours), so longer-term usage patterns and fatigue effects across full shifts remain unexplored. Despite these limitations, the findings still illustrate that embedding HSI principles can substantially improve IDS effectiveness and operator experience.

5. Implementation Framework

5.1. System Design

Designing cybersecurity frameworks for smart grid SCADA systems requires integrating human factors, automation, and ethical considerations to create robust and user-friendly systems [6]. Emphasizing the TOP Model in system architecture ensures that system integration considers all socio-technical components, fostering emergent properties and operational resilience [6]. HITLS, for instance, helps anticipate behaviors and unintended consequences in the interaction between human operators and machine agents [6].

Prioritizing context-driven human-machine interfaces is essential to address cognitive workload [25]. As such, systems design must implement adaptive, context-sensitive interfaces that reduce information overload and facilitate timely, accurate decision-making. Hence, neglecting such considerations reduces efficiency and lowers usability, as seen with high cognitive demands in certain security management systems [25].

Integrating automation and continuous feedback optimizes detection and human performance [25]. Effective IDS design leverages automation, such as continuous integration and rapid feedback loops, without excluding the human operator [25]. This combines the technical benefits of automation (e.g., higher software quality and release velocity) with the need for real-time, actionable alerts tailored to operator needs [25]. Embedding ethical oversight and continuous symbiotic maturity evaluations is crucial for balancing autonomous operations with user experience and ethical considerations.

Embedding computational efficiency through targeted ML architecture and feature engineering enhances real-time performance and minimizes resource consumption, a critical factor for SCADA systems [24]. Adopting lightweight, energy-efficient models and using feature engineering informed by HSI principles are key design considerations [24]. Explicitly modeling human error as a design consideration is vital, recognizing that many breaches originate from human factors. System design should integrate quantifiable metrics for operator usability, develop interfaces that reduce misinterpretation, and facilitate trade-offs between security and ease of use to improve overall reliability. Anticipating evolving threats and communication environments, such as wireless and IoT constraints (e.g., 5G coverage variability and adversarial impacts on embedded models), requires designing for resilience against disruptions and faults [26].

5.2. Operational Guidelines

Operational guidelines for implementing cybersecurity systems must focus on practical best practices, process optimization, and maintaining a balance between security and usability. Operationalizing continuous human-AI feedback loops is critical for reducing false positives and response fatigue. Establishing real-time, adaptive feedback mechanisms between operators and ML-based IDS directly supports operator learning and adaptive system behavior. Explainability-driven feedback, for example, enables operators to identify and correct misclassified attacks, improving alert quality and decision-making.

Mandating the integration of human factor metrics in deployment and ongoing assessment is crucial. Organizations must enforce the inclusion of operator usability scores, cognitive load assessments, and response time metrics as part of routine IDS evaluation and tuning. Given that human factors cause up to 27% of security breaches and that neglecting the human component undermines technical investments, these metrics are vital for assessing overall system effectiveness.

Aligning operational procedures with NIST Cybersecurity Framework Profiles provides a structure for tailored risk management. Practical deployment should systematically map SCADA IDS operational practices to the NIST Framework Core functions (Identify, Protect, Detect, Respond, and Recover), using tailored implementation tiers and profiles to address each organization's specific risk profile and maturity. Embedding ethical oversight and continuous symbiotic maturity evaluations in day-to-day operations is also recommended for sustaining responsive and adaptive cybersecurity frameworks [27].

Establishing cross-functional operator training and multidisciplinary incident response protocols is essential. Guidelines should mandate joint training for cyber operators, IT specialists, and organizational leaders on the ML-IDS's technical aspects and the human factors influencing alert management and system reliability. Empirical studies show that the effectiveness of security technologies ultimately depends on well-integrated human and organizational processes.

6. Conclusion

This study presented a human-centric machine learning intrusion detection system for Smart Grid SCADA environments, grounded in Human-Systems Integration (HSI) theory. By reducing FPR from 6.2% to 4.5% in XGBoost models and lowering operator cognitive workload (NASA-TLX: 39.4) while improving usability (SUS: 76.2), our human-centric pipeline aligns technical capabilities with real-world operational needs. As a result, combining performance metrics with operator-focused usability evaluations, the study demonstrates that embedding human factors into IDS architecture enhances both system trust and incident response. While advanced ML-based IDSs have achieved impressive detection accuracy on benchmark datasets, their effectiveness in real-world operational environments is significantly impacted by human factors, including cognitive overload from excessive alerts, the lack of model interpretability, and the persistent challenge of human error.

Key takeaways include: (1) ML models like Random Forest and XGBoost provide high accuracy when trained on curated SCADA datasets; (2) human-centered interfaces reduce false positives and cognitive burden; and (3) HITL simulations offer a practical method to validate IDS in real-world conditions.

Future research and development must prioritize reducing false positive rates, improving alert explainability, and developing dynamic alerting strategies for operator cognitive states. In addition, while we targeted widely adopted algorithms, exploring deep-learning and federated approaches remains future work. Operational guidelines should mandate the inclusion of human factor metrics, establish continuous human-AI feedback loops, and ensure cross-functional training to build resilient socio-technical systems capable of defending critical smart grid infrastructure against increasingly sophisticated cyber threats. Although limited by dataset realism and sample size, these findings offer a compelling case for embedding human factors into ML-IDS research and deployment.

Overall, this research to our knowledge, being the first to quantify human factors in SCADA IDS using HSI theory, contributes a novel, scalable, and evidence-based framework that reinforces the necessity of integrating human systems theory into cybersecurity tools for critical infrastructures.

Acknowledgement

I am deeply grateful to my advisor and mentor for their unwavering support, insightful feedback, and intellectual rigor, which have guided this research from its earliest stages. I also extend sincere thanks to the members of my review committee—their expertise and thoughtful critiques have continually challenged me to deepen my analysis and broaden my perspective. The stimulating conversations and collaborative spirit of my scholarly community have been invaluable, making this work both intellectually rewarding and personally fulfilling.

Above all, I owe endless gratitude to my family and my wife. Your unconditional belief in me, patience, and sacrifices have made this journey possible. This achievement belongs to each of you as much as it does to me.

References

[1] J. Marron, A. Gopstein, N. Bartol, and V. Feldman, "Cybersecurity Framework Smart Grid Profile (NIST Technical Note 2051)," National Institute of Standards and Technology, 2019.

[2] A. Shehod, "Ukraine power grid cyberattack and US susceptibility: Cybersecurity implications of smart grid advancements in the US (CISL# 2016-22)," Cybersecurity Interdisciplinary Systems Laboratory (CISL), Sloan School of Management, Massachusetts Institute of Technology, 2016.

[3] N. Sahani, R. Zhu, J.-H. Cho, and C.-C. Liu, "Machine learning-based intrusion detection for smart grid computing: A survey," *ACM Trans. Cyber-Phys. Syst.*, pp. 1–23, 2023.

[4] K. Dietz *et al.*, "The missing link in network intrusion detection: Taking AI/ML research efforts to users,"

IEEE Access, vol. 12, pp. 79815–79837, 2024.

[5] B. Naqvi, N. Clarke, and J. Porras, "Incorporating the human facet of security in developing systems and services," *Information and Computer Security*, pp. 1–23, 2020.

[6] G. A. Boy, "Human-Systems Integration," in *The Palgrave Encyclopedia of the Possible*, Springer, 2021, pp. 1–11.

[7] R. Mittu and W. F. Lawless, "Human factors in cybersecurity and the role for AI," Association for the Advancement of Artificial Intelligence, 2014.

[8] S. Katiforis, "Synchronized coevolution: A conceptual framework for sustaining a human-centered security culture in AI-driven environments," Thesis, Laurea University of Applied Sciences, 2024.

[9] Z. Huang, "Human-centric training and assessment for cyber situation awareness," Doctoral dissertation, University of Delaware, 2015.

[10] J. Kamsamrong *et al.*, "State of the art, trends and skill-gaps in cybersecurity in smart grids," Cybersecurity Curricula Recommendations for Smart Grids (CC-RSG), 2022.

[11] N. Turner *et al.*, "The role of human factors in delivering cyber security," Chartered Institute of Ergonomics & Human Factors, 2023.

[12] M. Aurangzeb *et al.*, "Enhancing cybersecurity in smart grids: Deep black box adversarial attacks and quantum voting ensemble models for blockchain privacy-preserving storage," *Energy Reports*, vol. 11, pp. 2493–2515, 2024.

[13] V. Agate, F. M. D'Anna, A. De Paola, P. Ferraro, G. Lo Re, and M. Morana, "A behavior-based intrusion detection system using ensemble learning techniques," in *ITASEC'22: Italian Conference on Cybersecurity*, CEUR Workshop Proceedings (CEUR-WS.org), 2022, pp. 1–10.

[14] S. Y. Diaba *et al.*, "SCADA securing system using deep learning to prevent cyber infiltration," *Neural Networks*, vol. 165, pp. 321–332, 2023.

[15] P. Prjevara and D. van de Wouw, "Improving Machine Learning based Intrusion and Anomaly Detection on SCADA and DCS using Case Specific Information," System and Network Engineering, 2018.

[16] S. Jamshidi, A. Nikanjam, K. W. Nafi, F. Khomh, and R. Rasta, "Application of deep reinforcement learning for intrusion detection in Internet of things: A systematic review," Polytechnique Montréal, 2024.

[17] H.-T. Vo, N. N. Thien, K. C. Mui, and P. P. Tien, "Securing networks: An in-depth analysis of intrusion detection using machine learning and model explanations," *IJACSA-International Journal of Advanced Computer Science and Applications*, vol. 15, no. 5, pp. 1436–1444, 2024.

[18] F. Piekert *et al.*, "Human factors-centric validation of a security management system in a linked critical infrastructures environment," in *Intelligent Human Systems Integration (IHSI 2025)*, 2025, vol. 160, pp. 416–430.

[19] R. Duraz, D. Espes, J. Francq, and S. Vaton, "Explainability-based metrics to help cyber operators find and correct misclassified cyberattacks," in *SAFE'23*, 2023.

[20] J. Yang, Y. Liu, and P. L. Morgan, "Human-machine interaction towards Industry 5.0: Human-centric smart manufacturing," *Digital Engineering*, vol. 2, no. 2, pp. 1–17, 2024.

[21] K. Kucuk, E. I. Yurteri, and B. Semiz, "Electroencephalography analysis frameworks for the driver fatigue problem: A benchmarking study," in *Proceedings of the 18th International Joint Conference on*

Biomedical Engineering Systems and Technologies (BIOSTEC 2025) - Volume 1, SCITEPRESS - Science and Technology Publications, Lda, 2025, pp. 829–836.

[22] S. Rajagopal, P. P. Kundapur, and K. S. Hareesha, "A predictive model for network intrusion detection using a stacking approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 3, pp. 2734–2741, 2020.

[23] J. S. Chavis, "Toward assurance and trust for the Internet of things," Doctoral dissertation, Johns Hopkins University, 2021.

[24] S. Jamshidi, K. W. Nafi, A. Nikanjam, and F. Khomh, "Evaluating machine learning-driven intrusion detection systems in IoT: Performance and energy consumption," Preprint submitted to Elsevier, 2024.

[25] O. M. Elazhary, "Exploring the socio-technical impact of continuous integration: Tools, practices, and humans," Doctoral dissertation, University of Victoria, 2021.

[26] F. Tu Zahra, Y. S. Bostanci, and M. Soyturk, "Security of Wireless IoT in Smart Manufacturing: Vulnerabilities and Countermeasures," in *Intelligent Secure Trustable Things*, Springer Nature, 2024, pp. 419–441.

[27] M. Saadallah, A. Shahim, and S. Khapova, "Optimizing AI and human expertise integration in cybersecurity: Enhancing operational efficiency and collaborative decision-making," *PriMera Scientific Engineering*, vol. 6, no. 2, pp. 03–20, 2025.