Supplementary Material: Insider Threat Detection using Machine Learning SLR

A Systematic Literature Review (SLR)

Publication Details:

- **Submitted to Journal:** IEEE Access
- **Submission Date:** February 12, 2026
- **Current Version:** February 12, 2026
- **DOI:** [10.1109/ACCESS.2026.XXXXXXX] (Pending)
- **Corresponding Author:** Qasim M. Alriyami (mohamedmuhanna@graduate.utm.my)

Note: this Supplementary material can also be accessed online on GitHub at: https://github.com/alriyami7777-alt/Insider-Threat-ML-SLR-Supplement

---

1. Overview

This repository contains the complete supplementary materials and "audit trail" for our Systematic Literature Review (SLR) on Insider Threat Detection using Machine Learning. This review synthesizes **82 primary studies** published between 2021 and 2025.

The repository provides full transparency into our methodology, including our search strategy, the PRISMA-2020 selection process, quality assessment scoring, and the AI-assisted data extraction process and pipeline.

2. Repository Structure

The repository is organized into five folders:

Folder Name	Primary Content	Format
01_Search_Strategy	Exact Boolean strings, database search logs, and raw record counts. RIS and CSV files from the databases	PDF RIS CSV
02_Study_Selection	PRISMA 2020 Flow Diagram and the Exclusion Log for the 16 rejected full-text papers.	PDF / CSV
03_Quality_Assessment	Detailed QA1–QA5 scoring for 98 candidate articles and scoring definitions.	PDF / Excel
04_Data_Extraction	Structured data extraction for RQ1–RQ5, including AI-assisted extraction methodology.	PDF
05_Visualizations	Bibliometric data, JCR quartile distributions, and keyword frequency datasets.	PDF / Excel

3. Transparency

To ensure the highest level of academic rigor, this repository documents two specialized workflows used in our study:

- **AI-Assisted Extraction (Folder 04):** We utilized **NotebookLM** for the initial extraction of technical parameters from the 82-paper. All AI-generated data was subsequently audited and manually verified by the authors.
- **Dual-Stream Bibliometric Pipeline (Folder 05):** Visualizations were generated through a hybrid pipeline combining qualitative thematic scans (NotebookLM) and quantitative metadata mapping (Zotero/Excel).

4. How to Use This Repository

- **To Verify Selection:** Open `02_Study_Selection/Exclusion_Log.pdf` to see the specific reasons (Secondary studies vs. QA failures) for paper rejection.
- **To Review Technical Data:** Consult the five PDFs in `04_Data_Extraction/` for detailed tables on preprocessing, model architectures, and performance metrics.
- **To Audit Quality:** See `03_Quality_Assessment/QA_Scoring_Master_List.xlsx` for the scoring used to reach the 3.5/5.0 inclusion threshold.

The repository also contains the PRISMA 2020 for Abstracts Checklist and PRISMA 2020 Checklist