Appendix A: Systematic Search Protocol

AI-Based Detection of Hedge Fund Fraud

Systematic Literature Review Methodology

SALSA Methodology for Systematic Review

Framework

SALSA methodology adapted from clinical research to accommodate interdisciplinary AI/ML applications in financial fraud detection.

Search

- Define search queries
- Select databases
- Execute searches
- Track results

AppraisaL

- Screen titles/abstracts
- Apply inclusion criteria
- Full-text review
- Quality assessment

Synthesis

- Extract key findings
- Organize by themes
- Identify patterns
- Note gaps

Analysis

- Compare methods
- Evaluate effectiveness
- Draw conclusions
- Future directions

Ensures comprehensive coverage of AI-based hedge fund fraud detection literature

Grant & Booth (2009). SALSA framework for systematic literature review.

Search Databases and Coverage

Database Selection

Five major academic databases ensure comprehensive coverage of computer science and finance literature.

Database	Coverage	Primary Focus
Scopus	Multidisciplinary abstracts and citations	Broad academic coverage
Web of Science	Citation database across sciences	High-impact journals
IEEE Xplore	Computer science and engineering	AI/ML methodology
SSRN	Social science research network	Finance working papers
Google Scholar	Supplementary coverage	Citation tracking, gray literature

Search Period:

- 2000–2025 (25 years)
- Emphasis on 2015–2025 (deep learning era)

Rationale:

- Captures modern AI/ML techniques
- Includes foundational statistical methods

Searc

conducted 2000–2025 with emphasis on deep learning period (2015–2025).

Boolean Query Structure

Query Design

Primary search query combines three conceptual blocks using Boolean operators to capture interdisciplinary literature.

Block 1: Financial Domain

- "hedge fund"
- "investment fund"
- "alternative investment"

OR operator

Block 2: Fraud Terms

- "fraud"
- "manipulation"
- "anomaly"

OR operator

Block 3: AI/ML Methods

- "machine learning"
- "artificial intelligence"
- "deep learning"
- "neural network"

OR operator

Primary Query:

(Block 1) **AND** (Block 2) **AND** (Block 3)

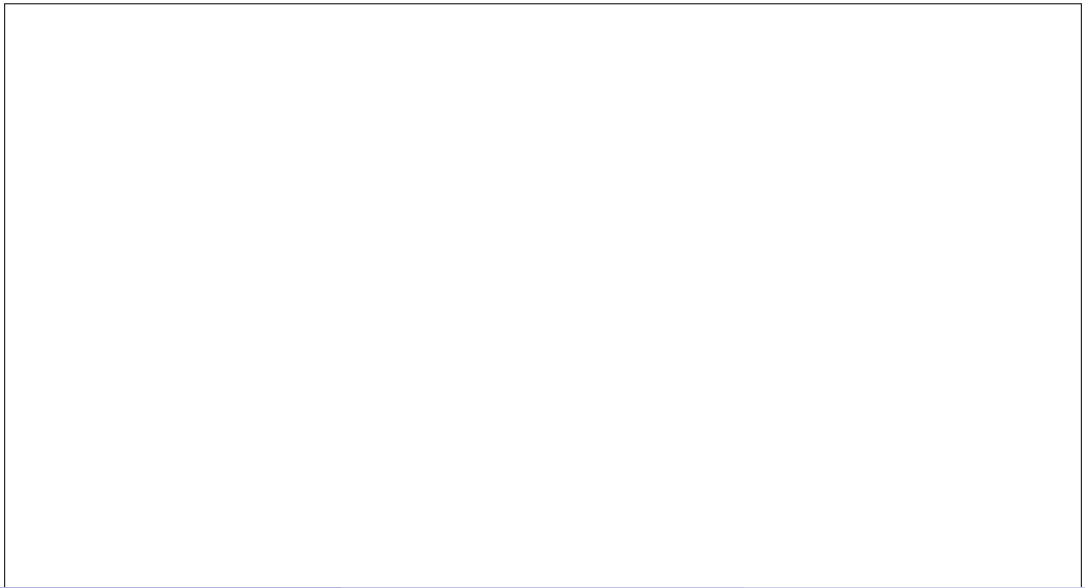
Secondary Query: Broader terms to capture related research

- ("fund" OR "portfolio") AND ("fraud detection" OR "anomaly detection") AND ("classification" OR "prediction")
- Captures mutual fund and general portfolio fraud studies

query designed to balance precision and recall across interdisciplinary literature.

Boole

Boolean Query Visualization



Inclusion and Exclusion Criteria

Inclusion Criteria

1. **Domain:** Fraud, manipulation, or anomaly detection in investment funds or alternative investments
2. **Methods:** Employed or substantively discussed AI/ML methods for detection, prediction, or classification
3. **Quality:** Peer-reviewed publications or widely cited preprints (minimum 20 citations)
4. **Language:** Published in English

Exclusion Criteria

1. **Pure Methodology:** Papers without financial application or validation
2. **Different Domain:** Studies focused exclusively on credit card, payment, or transaction-level fraud without fund-level components
3. **Duplicates:** Multiple publications reporting the same study across venues

Quality Threshold: Peer-reviewed publications or preprints with at least 20 citations ensures rigor and impact while capturing emerging research.

Crite

designed to ensure relevance, quality, and comprehensive coverage.

Screening Process and Results

Multi-Stage Screening

Two independent reviewers conducted screening with disagreements resolved through discussion and third-reviewer consultation.

Screening Stages:

1. **Database Search:** Initial queries across 5 databases
2. **Title/Abstract Screening:** Relevance assessment
3. **Full-Text Review:** Detailed examination against inclusion criteria
4. **Snowballing:** Manual review of highly cited paper references

Quality Assurance:

- Two independent reviewers
- Disagreement resolution protocol
- Third reviewer consultation as needed

Quantitative Results:

screening process: 500 → 120 → 80 + 25 = 105 papers.

Summary: Systematic Search Protocol

Comprehensive Methodology

SALSA-based systematic review ensures comprehensive coverage while maintaining rigorous inclusion standards.

Key Strengths:

- **Comprehensive:** 5 databases, 25-year coverage
- **Rigorous:** Independent dual review, quality thresholds
- **Interdisciplinary:** AI/ML + finance literature
- **Transparent:** Documented inclusion/exclusion criteria
- **Complete:** Snowballing captures additional relevant work

Final Corpus:

- **80 core papers** on AI/ML hedge fund fraud detection
- **25 contextual papers** on regulation, foundational methods
- **105 total references** providing comprehensive coverage
- Peer-reviewed publications and high-impact preprints
- Balanced coverage of methods and applications

Outcome: Systematic protocol ensures survey provides authoritative, comprehensive, and unbiased coverage of AI-based hedge fund fraud detection research.

SALS

methodology ensures comprehensive, rigorous, and transparent literature coverage.