

1. Search Strategy

We employed a reproducible search strategy to identify relevant studies on ML-based insider threat detection. The SLR targeted five of the most sought-after academic databases in the field. The five databases were Web of Science, Scopus, ScienceDirect, IEEE Xplore, and ACM Digital Library. The table below summarizes the search strings and the number of records retrieved from each database.

Database	Search String	Results
Web of Science	("insider threat") AND ("machine learning" OR "deep learning" OR "ensemble learning") AND ("model" OR "framework" OR "technique")	215
Scopus	TITLE-ABS-KEY ("insider threat" AND ("machine learning" OR "deep learning" OR "ensemble learning") AND ("model" OR "framework"))	343
ScienceDirect	"insider threat" AND ("machine learning" OR "deep learning" OR "ensemble learning") AND "detection"	902
IEEE Xplore	("insider threat" AND ("machine learning" OR "deep learning" OR "anomaly detection") AND ("model" OR "framework" OR "technique"))	439
ACM Digital Library	"query": ("insider threat" AND ("machine learning" OR "deep learning" OR "anomaly detection") AND ("model" OR "framework" OR "technique")) "filter": Article Type: Research Article, Survey, E-Publication Date: (01/01/2021 TO 12/31/2025)	50
Total Studies Identified		1949

2. Files with Raw Results

To further support the transparency of the systematic review process, the following table details the specific metadata for each search execution. This includes the exact dates on which the searches were performed and the corresponding export filenames stored within this directory. We provide this 'audit trail' to ensure that the search results can be independently verified and mapped back to the original database exports retrieved between December 2025 and January 2026.

Database	Date of Search	Result Count	Primary Export Filename	No of Files
Web of Science	Dec 20, 2025	215	WoS Original 215 Papers.xls	1
Scopus	Dec 23, 2025	343	scopus_export_Dec 23-2025...ris	1
ScienceDirect	Dec 23, 2025	902	ScienceDirect_citations_1766...ris *The files contain 389 records	4
IEEE Xplore	Dec 27, 2025	439	IEEE Xplore Citation RIS...ris **The Files contain 44 records	2
ACM Digital Library	*Jan 29, 2026	50	***acm Import 50.enw	1
Total Identified		1,946		

Notes:

* The 4 RIS files were only downloaded from ScienceDirect after applying the automatic filters (year and paper type filters). This resulted in 389 papers out of the original 902 records found using the detailed search string.

** The two RIS files were only downloaded from IEEE Explore after applying the automatic filters (year and paper type filters). This resulted in 44 papers out of the original 439 records found using the detailed search string.

***The search for the ACM Digital Library was re-executed on January 29, 2026, to ensure complete documentation of the initial filtration phase. During the preliminary search, all 50 records identified in this database were excluded in the first stage of filtration. Specifically, 43 of these sources did not meet the required publication window of 2021–2025. Furthermore, only four of the seven remaining papers were identified as peer-reviewed journal articles, and upon detailed "title" review, none were found to be relevant to the specific domain of machine learning-based insider threat detection. Consequently, no primary files were downloaded during the initial search in December 2025. The current repository reflects the secondary search conducted in January 2026 to verify the exclusion results.

3. Study Filtration and Selection Process

Following the initial identification of 1,946 records, a multistage filtration process was implemented to reach the final synthesis of 82 primary studies. This process adhered to the **PRISMA-2020** guidelines to ensure a systematic selection of peer-reviewed journal articles.

Table 3: Multi-Stage Filtration Results by Database

	Initial Results	Year 2021 2025	Journals Only			
	Stage 1 (full search Count)			Stage 2 (title Only)	Stage 4 (title and Abstract)	Stage 4 (Full Text)
Web of Science	215	149	118	58	40	36
Scopus	343	255	108	84	40	34
Science Direct	902	659	389	27	8	3
IEEE Explore	439	272	44	31	10	9
ACM Digital Library	50	43	4	0	0	0
Remaining Papers	1949	1378	663	200	98	82
Excluded	-571	-715	-463	-102	-16	

As illustrated in the filtration table, the screening process involved the following steps:

- **Stages 1 and 2:** Removing records outside the 2021–2025 range and non-journal publications, followed by screening of titles.
- **Stages 3 and 4:** Screening by Title and Abstract, followed by a full-text eligibility assessment.

The folder also contains the EXCEL file used to ensure the calculations above are accurate

The next folder (02_Study_Selection) gives more details on the screening and selection process following the PRISMA methodology