VARDHAMAN
COLLEGE OF ENGINEERING

(AUTONOMOUS)

Affiliated to **JNTUH**, Approved by **AICTE**, Accredited by **NAAC** with **A++** Grade, **ISO 9001:2015** Certified
Kacharam, Shamshabad, Hyderabad – 501218, Telangana, India

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING (AI&ML)

Literature Survey Report

Team number: 22

Title of the project: Anomaly Detection in Host Systems Using Random Forest-Based HIDS

S.No	Methodology Used	Dataset Used	Performance Metrics	Pros	Cons
[1]	Multi-level Random Forest with Fuzzy Inference System	NSL-KDD dataset	Accuracy: 98.2%, Precision: 97.9%, Recall: 98.0%	Robust hybrid model (BERT-LSTM), high accuracy, low overfitting	Resource-intensive, complexity in deployment
[2]	Systematic review of unsupervised learning methods (clustering, deep learning, statistical) for anomaly detection in traffic flows	NSL-KDD, UNSW-NB15, CICIDS2017, BoT-IoT	No experimental results; metrics summarized from literature (e.g., accuracy, DR, FPR)	Highlights key datasets, unsupervised methods, trends, and gaps in literature	Does not propose or test a new model; depends on third-party results
[3]	Eigentraces-based feature extraction + One-Class SVM and One-Class Nearest Neighbor (OC-NN) for anomaly detection in system calls	System Call Trace Dataset from UNM	One-Class SVM Accuracy: 95.3%, OC-NN Accuracy: 94.6%	Lightweight, effective on host-level syscall behavior, good for zero-day detection	Only works with syscall data, sensitive to noise and system variability

[4]	Enhanced Random Forest with optimized feature selectio	CICIDS2017	Accuracy: 99.1%, F1-score: 98.8%	Improved model efficiency, robust to noise, good generalization	May underperform on unseen attack types
[5]	Genetic Algorithm–optimized feature selection + Random Forest (GA-RF) ensemble	NSL-KDD (148,517 instances, 41 features) UNSW_2018_IoT_Botnet	Accuracy: 99.999%; Precision: 100%; Recall: 100%	Extremely high detection accuracy Very low false positives Suitable for IoMT environments with critical reliability needs	May not generalize beyond tested datasets Potential overfitting concerns Genetic Algorithm adds computational complexity

[1] B. Awotunde et al., “A Multi-level Random Forest Model-Based Intrusion Detection Using Fuzzy Inference System”, Int. J. Comput. Intell. Syst. (2023).

[2] Alberto Miguel-Diez, Adri´an Campazas-Vega, Claudia ,Alvarez-Aparicio, Gonzalo,Esteban-Costales, and Angel Manuel Guerrero-Higuer,“ A systematic literature review of methods and datasets for anomaly intrusion detection, Computers & Security” (2023).

[3] Ehsan Aghaei 1, Gursel Serpen2 ,”Host-based anomaly detection using Eigentraces feature extraction and one-class classification on system call trace data”, Sensors (2024).

[4] Caiwu Lu Yunxiang Cao and Zebin Wang,” Research on Intrusion Detection Based on an Enhanced Random Forest Algorithm”(2024).

[5] Monire Norouzi , Zeynep Gürka¸s-Aydın , Özgür Can Turna , Mehmet Yavuz Ya ˘gci , Muhammed Ali Aydin and Alireza Souri ,” A Hybrid Genetic Algorithm-Based Random Forest Model for Intrusion Detection Approach in Internet of Medical Things”(2023)

Signature of the Supervisor Signature of Coordinator Signature of HOD