B.M.S. COLLEGE OF ENGINEERING, BENGALURU

Autonomous Institute, Affiliated to VTU

An AAT Report on
on

"IDRandom-Forest: Advanced Random Forest for Real-Time Intrusion Detection"

Submitted in fulfilment for the award of degree of
Master of Technology
In

Computer Science and Engineering

Submitted by

Revanth L (1BM25SCS007)

Under the Guidance of

Dr. Umadevi V
Professor, BMSCE

B.M.S. COLLEGE OF ENGINEERING
(Autonomous Institution under VTU)
BENGALURU-56001
9 2025-2026

B.M.S. College of Engineering,

Department of Computer Science and Engineering

DECLARATION

This is to declare that the AAT work entitled "**IDRandom-Forest: Advanced Random Forest for Real-Time Intrusion Detection"** carried out by **Revanth L (1BM25SCS007)**, is a part of the course Artificial Intelligence (MCS101) of 1st Semester M.Tech (CSE), Department of Computer Science and Engineering, B.M.S. College of Engineering, Bangalore. This AAT work has been carried out under the guidance of Dr. Umadevi V during the academic semester Nov 2025 - Feb 2026. I declare that this work is original and has not been submitted elsewhere for any other degree or award.

Signature of the Candidate

Revanth L (1BM25SCS007)

B. M. S. COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

Certified that the AAT entitled **"IDRandom-Forest: Advanced Random Forest for Real-Time Intrusion Detection"** is a bonafide work carried out by **Revanth L (1BM25SCS007)**, in partial fulfillment for the award of Master of Technology in Computer Science and Engineering of the Visvesvaraya Technological University, Belagavi, during the academic year 2025-2026. This report satisfies the academic requirements prescribed for the said degree.

Guide

Dr. Umadevi V
Professor
Department of CSE
BMSCE

Table of Contents

1. Introduction

The risk of cyberattacks has entirely increased because of the fast growth of computer networks and internet-based services. Network security is a important issue since businesses today depend more on digital infrastructure for online services, data storage, and communication. Network traffic is monitored by intrusion detection systems(IDS), which finds doubtful activity that could lead to security breaches. Conventional security techniques, like firewalls and antivirus software, are frequently insufficient because they concentrate mainly on known attacks and may miss new threats.

Machine learning and Deep learning techniques for intrusion recognition techniques have gained importance because of the capability to learn patterns from network data and detect anomalies.

2. Problem Statement

In recent days cyber attacks have increased. The privacy of the network has become a critical issue. For the real time Intrusion detection the present machine learning methods such as RF-SVM are characterized by low detection accuracy, At the same time, techniques using complex models such as recurrent neural neural network and transformer-based deep learning are facing challenges of high latency rate in detecting intrusion in real-time traffic.

IDRandom-Forest for real-time intrusion recognition have increased high accuracy and decreased latency rate in recognizing intrusion in real-time traffic. Systems finding difficulty to find malicious activity in real time as complexity of network traffic increased. The ability of many intrusion detection systems to identify new and unknown threats is limited by their reliance on pre-established rules or signatures. Because anomaly-based systems frequently struggle to differentiate between malicious activity and typical network behavior, there are so many false alarms. This affects system reliability and adds to the workload of security analysts.

3. Objectives of the Paper

The main aim is to study and implement a effective intrusion detection technique that have ability to identify malicious network activity with high accuracy and computational efficiency.

The major contributions of the study are as follows:

- A stratified sampling–enhanced feature weighting technique is utilized in the random forest.
- To select the optimal sun-ensemble a window-based accuracy technique is used in the random forest method.
- Random forest with will give better accuracy, use less computational resources, have less latency and testing time on dataset UNSW-NB15

4. Datasets Description

Class	Training	Testing	Total
Normal	56,000	37,000	93,000
Attack	1,19,341	45,332	1,64,673
Total	1,75,341	82,332	2,57,673

The UNSW-NB15 dataset was generated in Cyber Range Laboratory at UNSW Canberra using IXIA PerfectStorm tool. This environment was designed to replicate realistic network conditions by combining normal user activities with modern attack scenarios. To collect the raw network traffic data, the Tcpdump tool was used, resulting in approximately 100 GB of packet capture(PCAP) files. These captured packets represent a composition of legitimate network behavior and malicious activities noticed in real-world environments

5. Tools and Technology Used for Implementation

Operating System: Windows

Code Editor: Cursor IDE

Programming Language: Python

Frameworks: Scikit-learn

Libraries:

- Numpy
- Pandas
- Matplotlib
- Scikit-learn

Dataset: UNSW-NB

6. GitHub Link of the Code

https://github.com/Revanth-1707/Revanth-1707-IDRandom-Forest-Advanced-Random-Forest-for-Real-Time-Intrusion-Detection.git

7. Methodology

The IDRandom-Forest approach is a three-step process for building a efficient intrusion detection model. Initially, feature importance is analyzed using the Gini Index, which helps in grouping the features into strong and weak categories according to their contribution. In the second step, set of decision trees is generated using bootstrap sampling, where different subsets of features are selected to increase diversity among the trees. In the final step, an Accuracy Sliding Window (ASW) based pruning method eliminate trees that are not significantly ihelp to mprove performance. This process results in a smaller and more effective ensemble, while maintaining high detection accuracy.

7.1 Overall System Architecture

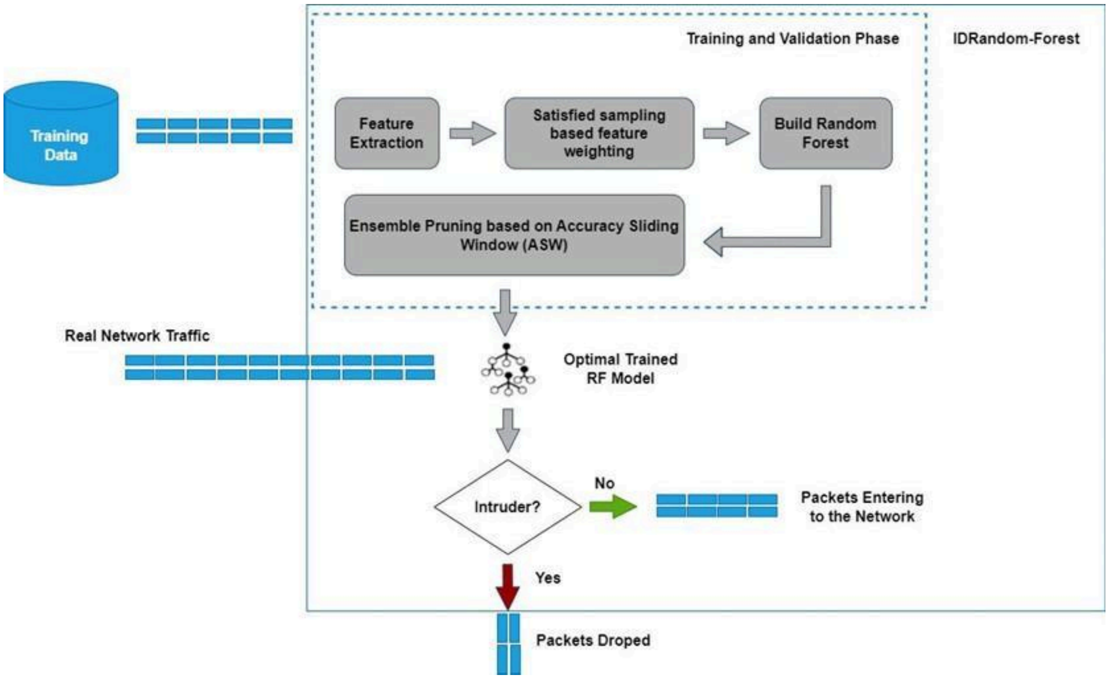

Training and Validation Phase
IDRandom-Forest
Training Data
Feature Extraction
Satisfied sampling based feature weighting
Build Random Forest
Ensemble Pruning based on Accuracy Sliding Window (ASW)
Real Network Traffic
Optimal Trained RF Model
Intruder?
No
Packets Entering to the Network
Yes
Packets Droped

Figure 7.1: Framework of the proposed IDRandom-Forest classifier.

7.2 Stratified Feature Weighting

To find how much each attribute contributes to the identification of intrusions, feature importance is examined at this stage. Each feature's importance is calculated through the Gini index to determine the feature's capacity to differentiate between regular and attack traffic. Features are categorized as strong or weak according to these values.

7.3 Building Decision Tree

Bootstrap sampling is used to extract a different set of dataset for each decision tree. Features are selected from the strong and weak feature groups created in the above step for tree construction. Helps to reduce overfitting and generalize will on new network threats.

7.4 Pruning Decision Trees

Accuracy sliding window mechanism used to prune the ensemble of decision trees according to how much each tree is contributing to the overall performance. Trees whose performance is less are removed. When a perfect accuracy to ensemble size ratio is reached, the pruning process is repeated

8. Results Obtained

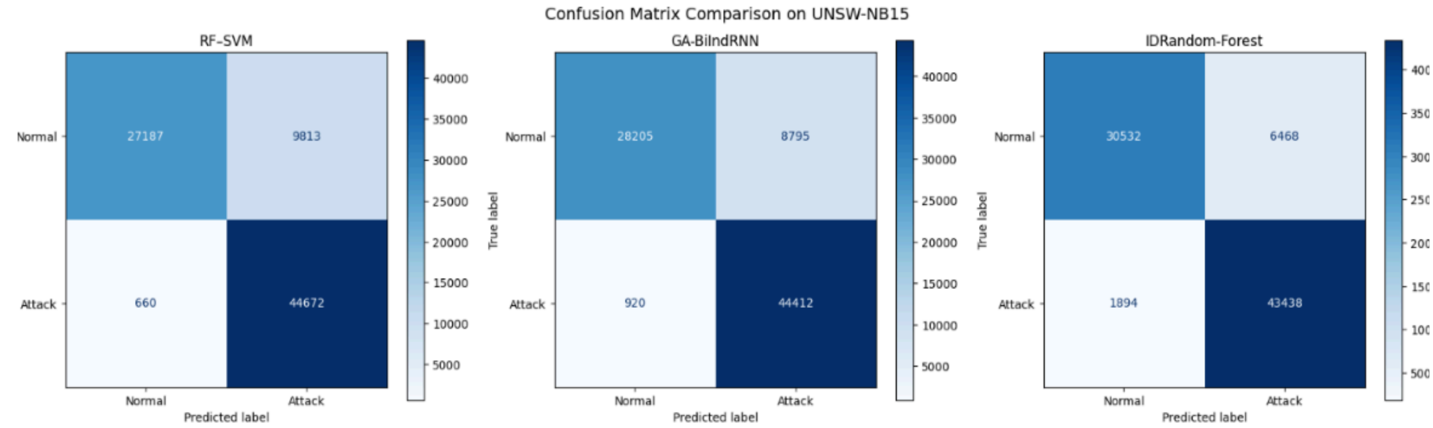
The UNSW-NB15 testing data utilized to assess efficiency of the proposed IDRandom-Forest model. The experimental findings tells that the model achieved high degree of accuracy in differentiating between normal and attack traffic; strong detection rate was observed, indicating majority of attack instances were accurately detected, but the false alarm rate stayed comparatively low, the precision and F1-score values further confirmed model's reliability in handling intrusion detection tasks; overall, results tells proposed approach is effective and suitable for real-time intrusion detection, offering good balance between accuracy and efficiency.

8.1 Model Performance Evaluation

	Accuracy	**Precision**	**Recall**	**F1-score**
RF-IDRF	90.04	91	90	90
SVM	87.28	90	86	87
BiIndRNN	88.24	90	87	88

Figure 8.1: Classification Repo

8.2 Confusion Matrix



8.3 ROC curve

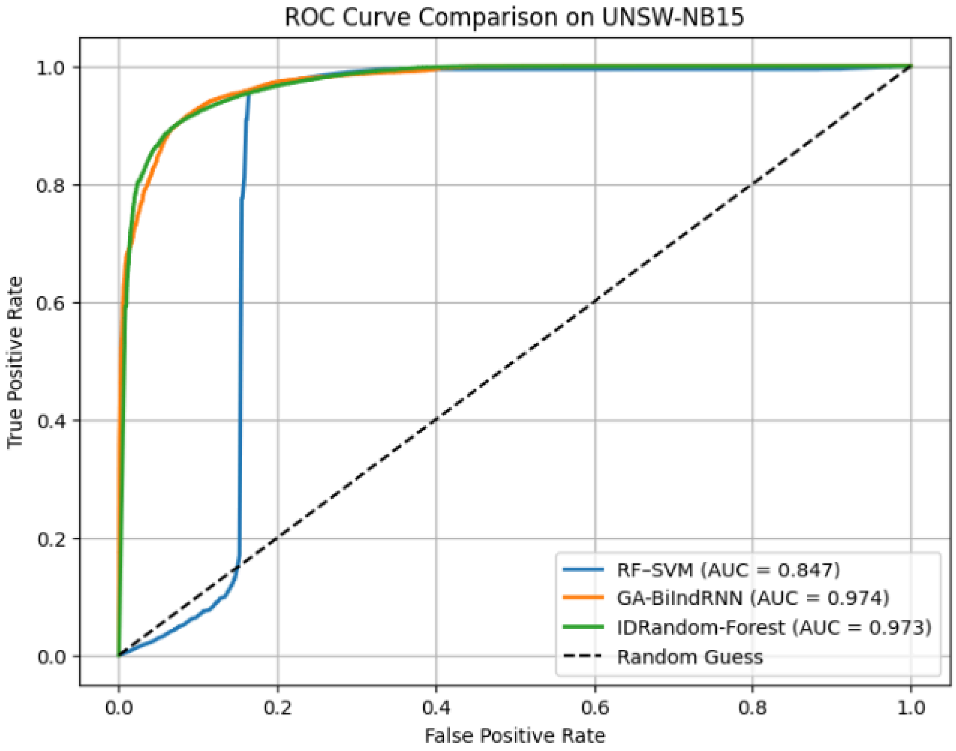
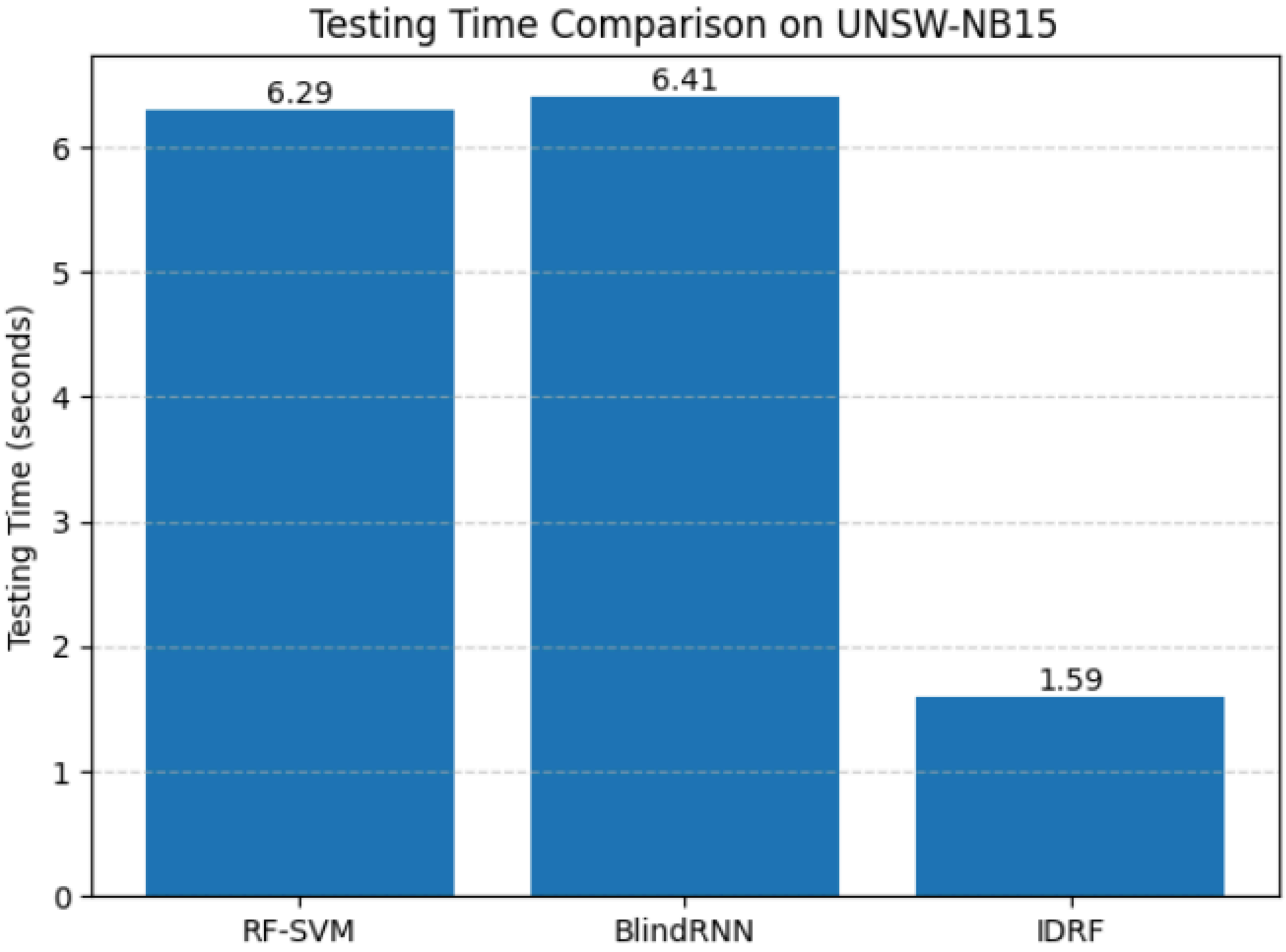


Figure 8.4: ROC curve

8.4 Testing Time Comparison



8.5 Discussion

The new mechanism like Accuracy Sliding Window (ASW) and Stratified feature weighting helps to remove weak trees that do not contribute positively to overall model performance and make sure every decision tree focuses on relevant attributes. The combination of these both mechanisms allows the model generalize well on new data or new types of attacks and avoid overfitting, which is important for real-time intrusion detection systems.

The clear performance variations are observed when comparing IDRF with machine learning methods such as RF-SVM, which means accuracy of RF-SVM is less compared to IDRF. BiIndRNN shows competitive performance however, it requires high computational resources. Results show that BiIndRNN takes high testing time compared to IDRandom-Forest

9. Learning Outcome

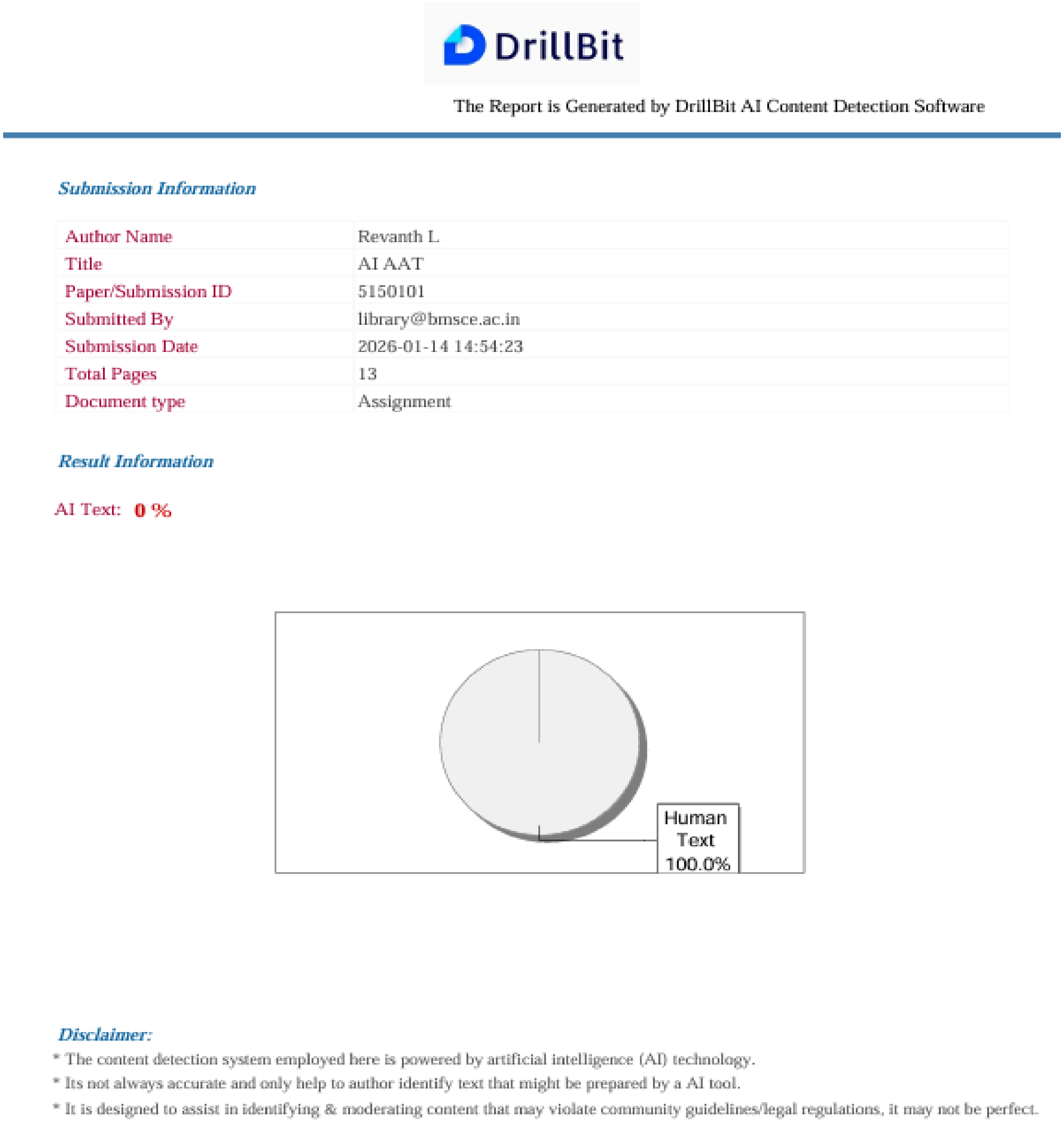
This work provided clear understanding of intrusion recognition systems and their significance for network security. IDRandom-Forest model made easier to understand how ensemble learning methods can enhance detection performance when compared to single classifiers. I also learnt handling real-world network traffic data and understanding different cyberattacks by working with the UNSW-NB15 dataset.

Developed practical skills in data preprocessing, model training, and performance evaluation using common metrics. I have a better understanding of how to increase the model's efficiency without reducing accuracy. The project helped me to improve my knowledge in developing, putting into practice, and evaluating intrusion detection models in a real-world context.

10. Paper Reference

1. M. Azhar, S. Perveen, A. Iqbal and B. Lee, "IDRandom-Forest: Advanced Random Forest for Real-Time Intrusion Detection," in IEEE Access, vol. 12, pp. 113842-113854, 2024, doi: 10.1109/ACCESS.2024.3443408.

11. Screenshot of Similarity & AI Generated report



DrillBit

The Report is Generated by DrillBit AI Content Detection Software

Submission Information

Author Name	Revanth L
Title	AI AAT
Paper/Submission ID	5150101
Submitted By	library@bmsce.ac.in
Submission Date	2026-01-14 14:54:23
Total Pages	13
Document type	Assignment

Result Information

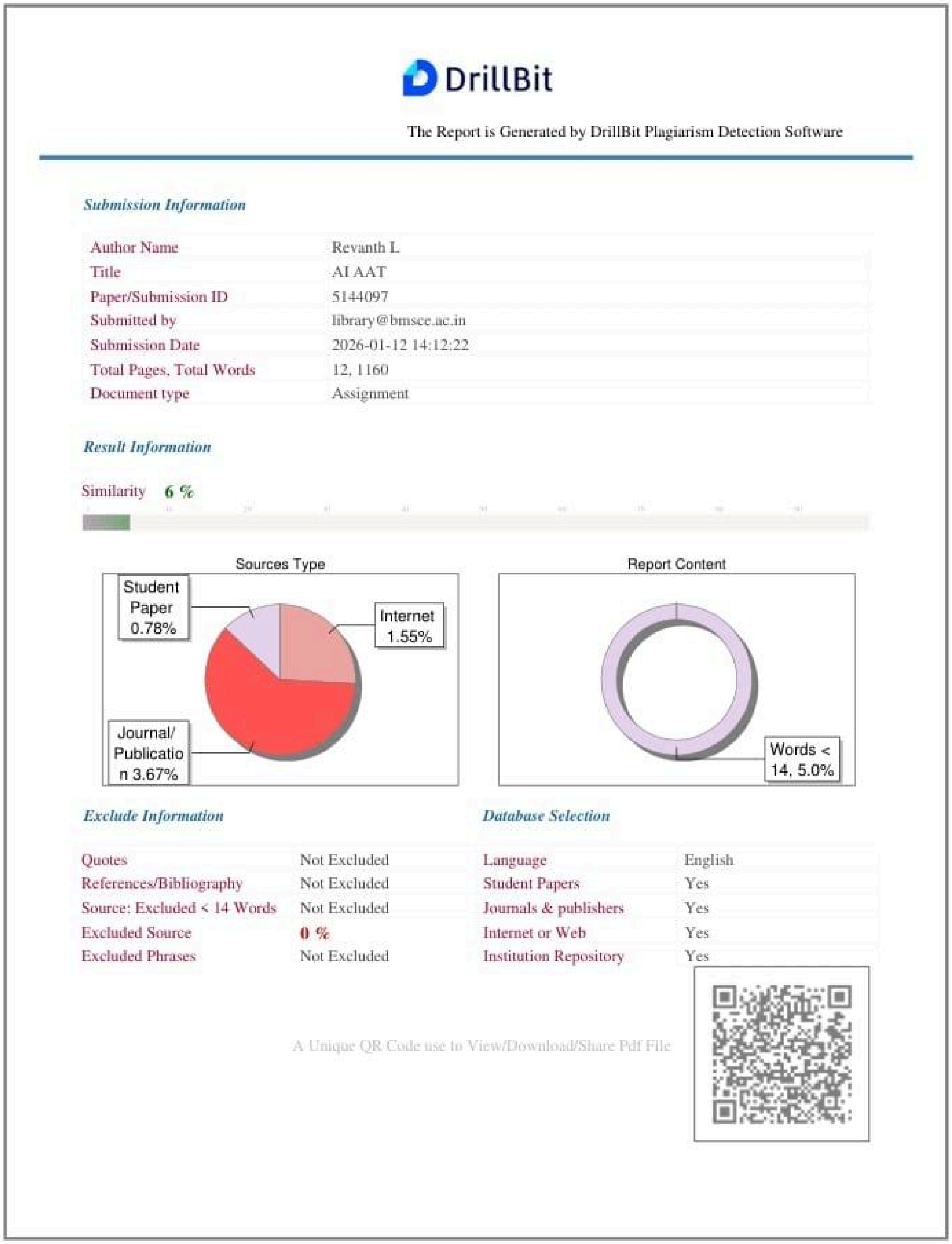
AI Text: **0 %**

Disclaimer:

* The content detection system employed here is powered by artificial intelligence (AI) technology.

* Its not always accurate and only help to author identify text that might be prepared by a AI tool.

* It is designed to assist in identifying & moderating content that may violate community guidelines/legal regulations, it may not be perfect.



DrillBit

The Report is Generated by DrillBit Plagiarism Detection Software

Submission Information

Author Name	Revanth L
Title	AI AAT
Paper/Submission ID	5144097
Submitted by	library@bmsce.ac.in
Submission Date	2026-01-12 14:12:22
Total Pages, Total Words	12, 1160
Document type	Assignment

Result Information

Similarity **6 %**

Sources Type

Report Content

Exclude Information

Quotes	Not Excluded
References/Bibliography	Not Excluded
Source: Excluded < 14 Words	Not Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

Database Selection

Language	English
Student Papers	Yes
Journals & publishers	Yes
Internet or Web	Yes
Institution Repository	Yes

A Unique QR Code use to View/Download/Share Pdf File