System for Identifying Human Activities and Detecting Suspicious Behaviours

[1]Anandi Bole, [2]Sweta Kale, [3]Priyanka Garje, [4]Aditya Gawade,[5]Surabhi Sharma

[1]Student, [2]Faculty,[3]Student,[4]Student,[5]Student

[1]Department of Information Technology,

RMD Sinhgad School of Engineering, Pune, India

Abstract: Suspicious activity identification from surveillance video is an effective research area of image processing and computer vision. Detecting suspicious activities is significant for maintaining the security of organizations and communities. Surveillance cameras are mostly used in public areas to monitor and secure safety. It is difficult to observe public places continuously hence intelligent video surveillance is needed that can detect human activity in real-time and classify them as non-suspicious & suspicious activities. By employing a stationary camera and integrating state-of-the-art machine learning algorithms—including Logistic Regression, Ridge Classifier, Random Forest, and Gradient Boosting—the system achieves real-time identification of diverse activities. The dataset, constructed from human body key points extracted through video analysis, demonstrates the system's efficacy in distinguishing between normal and anomalous behaviours. With potential applications spanning surveillance, healthcare, and public safety, this research underscores the system's capability to deliver timely alerts, thereby enhancing security protocols and minimizing reliance on manual monitoring.

Keywords- Suspicious Activity, Non-Suspicious Activity Logistic Regression, Ridge Classifier, Random Forest, and Gradient Boosting.

I. Introduction

In today's technologically advanced world, CCTV surveillance has emerged as one of the most critical and impactful security measures for any premises. Widely deployed in hospitals, universities, shopping malls, and other public spaces but traditional CCTV systems record footage but do not analyse it in real time. Suspicious activities are often identified only after the fact, during manual review, which can lead to delayed responses. Activity Recognition Systems leverage machine learning and computer vision to analyse video feeds in real time, enabling immediate detection of suspicious behaviours.

CCTV Cameras are reactive by nature, as they rely on post-event analysis. Suspicious activities are often identified only after an incident has occurred. In Activity Recognition Systems, by using machine learning algorithms, these systems can identify patterns and anomalies indicative of suspicious behaviour, enabling proactive measures to mitigate risks before they materialize. In CCTV Cameras, Operators must manually scan through footage to identify events, which can be time-consuming and inefficient. Activity Recognition Systems provide actionable insights by highlighting specific events or behaviours, enabling security personnel to focus on critical situations rather than sifting through hours of footage.

Continuous monitoring of public spaces presents significant challenges, necessitating the adoption of intelligent video surveillance systems capable of real-time human activity detection and classification into suspicious and non-suspicious categories. Leveraging a stationary camera, this advanced system integrates cutting-edge machine learning algorithms—such as Logistic Regression, Ridge Classifier, Random Forest, and Gradient Boosting—to enable accurate and efficient activity recognition, ensuring robust real-time surveillance capabilities.

II. Literature Review

The related work suggests different approaches for detecting human behaviours from video. The objective of the works was to detect any abnormal or suspicious events in video surveillance. Video surveillance is crucial for indoor and outdoor security. The system's components, such as behaviour recognition, can classify activities as normal or suspicious in real-time [01, 02].

[1] The evolution of suspicious activity detection in surveillance videos has shifted from traditional computer vision techniques to advanced deep learning methods. Early approaches, such as background subtraction, motion analysis, optical flow, and statistical methods, relied on manual feature extraction and were limited by high computational costs, sensitivity to dynamic environments, and poor scalability. These methods struggled with complex scenarios like crowded areas and subtle anomalies. Deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), has addressed these limitations by automating feature extraction and capturing temporal dependencies. CNNs extract spatial features from video frames, while RNNs, such as Long Short-Term Memory (LSTM) networks, analyse temporal sequences. Combined CNN-LSTM architectures have emerged as a powerful solution for detecting suspicious activities. The proposed system by Amrutha et al. leverages a CNN-LSTM model, using VGG16 for feature extraction and LSTM for classifying activities as "suspicious" (e.g., fighting, fainting) or "normal" (e.g., walking, running). This aligns with current trends in deep learning, offering a robust, automated, and scalable approach to suspicious activity detection in surveillance systems.

[2] Human activity recognition (HAR) has gained prominence in fields like healthcare, robotics, and surveillance. Traditional radar-based HAR systems face challenges, particularly when human movement is perpendicular to the radar's line of sight, limiting their effectiveness. This paper introduces a novel distributed MIMO radar system to achieve direction-independent HAR, addressing these limitations. Early radar-based HAR systems relied on hand-crafted features, such as time-frequency representations and statistical measures, which required domain

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

expertise and struggled with complex activities. Recent advancements in deep learning, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have enabled more robust HAR systems by automating feature extraction and classification from raw radar data. The proposed approach uses a distributed MIMO radar system with two independent subsystems, offering multi-perspective observation of human activities. Key contributions include direction-independent HAR, deep learning-based feature extraction and classification using a DCNN, and feature fusion to enhance accuracy. Experiments demonstrated a classification accuracy of 98.52% for activities like falling, walking, standing, sitting, and picking up an object, showcasing the system's effectiveness.

[3]Human-Object Interaction (HOI) analysis is a critical area in computer vision with applications in surveillance, human-computer interaction, and autonomous systems. This survey explores state-of-the-art techniques and challenges in HOI analysis, focusing on advancements and the method proposed by Israr Akhter et al. Early approaches relied on handcrafted features like HOG, SIFT, and motion cues, combined with traditional classifiers such as SVM or HMM. However, these methods struggled with complex interactions and variations in appearance and pose. Deep learning has revolutionized HOI analysis, with techniques like Two-Stream CNNs (fusing spatial and temporal features), Spatial-Temporal Graph Neural Networks (modelling interactions as graphs), and Transformer-based models (capturing long-range dependencies) leading the way. Key challenges include occlusions, pose variations, complex interactions, and recognizing unseen interactions. The proposed method by Akhter et al. combines traditional and deep learning techniques, featuring preprocessing (e.g., median filtering), human silhouette extraction (using GMM and superpixel models), robust human and object detection, feature extraction (e.g., R-transform, angular movement), and classification (using SGD and RBM). This hybrid approach addresses some challenges but requires further comparison with state-of-the-art methods to fully assess its contributions.

[4] The author [4] discusses how surveillance videos can capture realistic anomalies and proposes a new method for learning anomalies by using both normal and anomalous videos. To avoid manually labelling each abnormal segment, the authors suggest using a deep multiple instances ranking framework that leverages weakly labelled training videos. They consider normal and anomalous videos as bags and video segments as instances, and use this approach to automatically learn a deep anomaly ranking model that can predict high anomaly scores for anomalous video segments. They also introduce a new large-scale dataset of 1900 surveillance videos, containing 13 types of realistic anomalies as well as normal activities. This dataset can be used for general anomaly detection or recognizing each of the 13 anomalous activities. The authors' method achieves significant improvement on anomaly detection performance compared to previous methods, and the dataset is challenging and offers opportunities for future research [4].

III. SYSTEM ARCHITECTURE

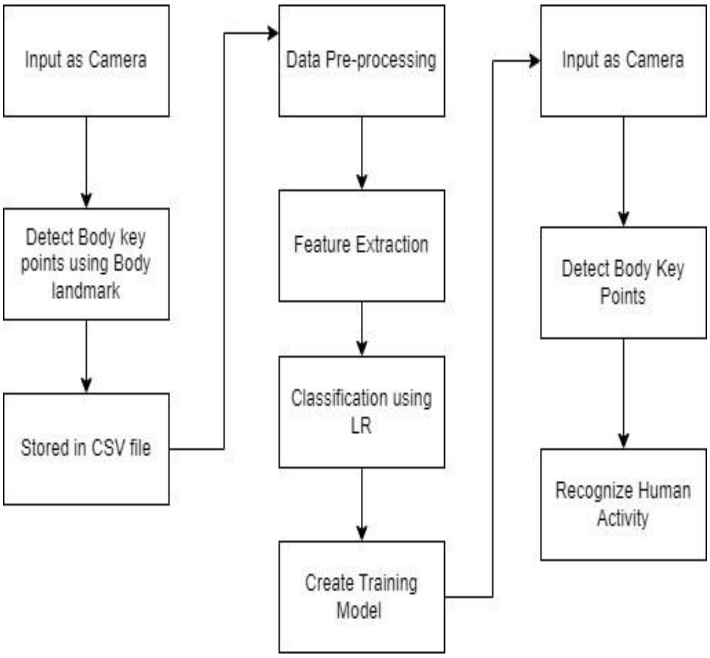

Fig.1: System Architecture

The system is designed for real-time human activity recognition and suspicious behaviour detection using video input from a stationary camera. It begins by capturing video footage, detecting key body points (e.g., joints), and storing this data in a structured CSV file. The data is then pre-processed to clean and normalize it, followed by feature extraction to identify meaningful movement patterns. A Logistic Regression (LR) model is trained on this data to classify activities. Once trained, the system can recognize and classify human activities in real-time, enabling applications in surveillance, security, and public safety.

1. Input from Camera: The system starts by capturing continuous video footage from a stationary camera. This video serves as the primary input for analysing human movements. The camera provides real-time visual data, which is essential for tracking and understanding human activities within the frame. High-quality video input ensures accurate detection and analysis of movements, forming the foundation for the entire system.
2. Detect Body Key Points Using Body Landmark Detection: Advanced algorithms (e.g., OpenPose, Media Pipe) are used to identify key points on the human body, such as joints (elbows, knees, shoulders) and other landmarks. These points represent the skeletal structure of individuals in the video. Detecting body key points allows the system to track posture and movement, enabling a detailed understanding of human actions. Accurate key point detection is critical for recognizing complex activities and distinguishing between normal and suspicious behaviours.
3. Store Data in CSV File:The coordinates of the detected body key points, along with metadata (e.g., timestamps, confidence scores), are stored in a structured CSV file. Storing data in a CSV file provides an organized and accessible format for further processing and analysis. A structured dataset facilitates efficient data retrieval, labelling, and preparation for machine learning tasks, ensuring scalability and reproducibility.

4. Data Pre-processing: The raw data undergoes cleaning, normalization, and transformation. This includes filtering out irrelevant frames, aligning data points, and handling missing or

noisy data. Pre-processing ensures the data is consistent, reliable, and ready for feature extraction and model training. Clean and normalized data improves the accuracy and efficiency of the system, reducing the risk of errors during analysis.
5. Feature Extraction: The system extracts meaningful features from the body key points, such as: Angles between joints: Captures the posture and orientation of body parts. Distances between key points: Measures the spatial relationship between body parts. Motion trajectories: Tracks the movement of key points over time. These features represent the essential aspects of human movement, enabling the system to differentiate between various activities. Feature extraction transforms raw data into actionable insights, making it easier for the model to learn and classify activities.
6. Classification Using Logistic Regression (LR): The extracted features are fed into a Logistic Regression model, a supervised machine learning algorithm. LR predicts the probability of each activity class (e.g., walking, sitting, waving) based on the input features. LR classifies activities by analysing patterns in the extracted features, assigning labels to movements in real-time. Logistic Regression is a simple yet effective algorithm for binary and multi-class classification, providing balance between accuracy and computational efficiency.
7. Create Training Model: Using the labelled dataset and extracted features, the system trains a machine learning model. The model learns to recognize patterns and correlations between features and activity labels. Training the model enables it to generalize from the training data and accurately classify new, unseen data. A well-trained model is essential for reliable and accurate activity recognition, ensuring the system performs effectively in real-world scenarios.
8.Recognize Human Activity in Real-Time: Once trained, the system analyses live video input, detects body key points, extracts features, and classifies activities in real-time. It compares detected movements with learned patterns to identify specific actions. Real-time activity recognition enables immediate detection of normal and suspicious behaviours, providing actionable insights for security and surveillance. Real-time analysis ensures timely responses to potential threats, enhancing the system's utility in practical applications.

IV. METHODOLOGY

The development of a system for detecting suspicious human activities is a complex, multi-stage process that integrates advanced machine learning techniques with robust data processing pipelines. The system is designed to analyze human behaviour through video or sensor data, identify key patterns, and classify activities as either normal or suspicious. This involves a series of well-defined steps, each critical to ensuring the system's accuracy, efficiency, and scalability. Below, we outline the methodology in detail, emphasizing the rationale behind each step and the choice of algorithms.

1. Data Acquisition

The process begins with **data acquisition**, where raw data is collected using stationary cameras, wearable sensors, or other IoT devices. This data serves as the foundation for all subsequent analysis. The quality and diversity of the dataset are paramount, as they directly influence the system's ability to generalize to real-world scenarios. To ensure robustness, the dataset includes:

- **Normal activities** (e.g., walking, sitting, running).
- **Suspicious behaviors** (e.g., loitering, sudden movements, aggressive gestures).
- **Varied environmental conditions** (e.g., lighting, occlusion, background noise).

The dataset is collected from multiple individuals to account for differences in posture, speed, and interaction with the environment. This diversity ensures the system can handle real-world variability and perform reliably across different settings.

2. Feature Extraction

Once the raw data is acquired, the next step is **feature extraction**, where meaningful attributes are derived to represent human movements. These features are critical for distinguishing between normal and suspicious activities. The feature extraction process involves:

- **Spatial metrics**: Joint angles, distances between body parts, and posture analysis.
- **Temporal metrics**: Speed, acceleration, and motion trajectories.
- **Frequency-domain transformations**: Periodic patterns in movement, such as gait analysis.

Advanced computer vision techniques like **OpenPose** and **MediaPipe** are used to detect body key points (e.g., elbows, knees, shoulders) from video data. These key points are then used to calculate features that describe posture, motion trajectories, and other aspects of human behaviour. The quality and relevance of these features directly influence the performance of the machine learning models.

3. Model Training and Classification

The extracted features are fed into machine learning models for training and classification. The choice of algorithms depends on the complexity of the activities, the need for real-time processing, and the availability of computational resources. Below, we discuss the algorithms used and their respective strengths:

1. Logistic Regression: A simple yet powerful algorithm for basic activity recognition tasks. Models the relationship between input features and activity labels. Suitable for classifying activities like walking, sitting, or running. Limitations: Its linear nature limits its ability to capture complex patterns.

2. Random Forest: An ensemble learning method that builds multiple decision trees and combines their outputs. Effective at handling complex activities involving multiple body parts. Identifies the most important features for classification, improving accuracy and robustness.

3.Gradient Boosting: Sequentially improves model performance by minimizing errors at each step. Excels at capturing subtle differences in similar activities. Well-suited for detecting sequential patterns in suspicious behaviour.

4.Ridge Classifier: Incorporates regularization to prevent overfitting and handle multicollinearity among features. Ideal for applications where data quality may vary. Reduces the impact of noise and improves the model's ability to generalize to new data.

5. Advanced Techniques: **Hidden Markov Models (HMMs)**: Represent activities as sequences of states, making them suitable for tasks involving a series of actions (e.g., walking followed by running).**Conditional Random Fields (CRFs)**: Model the conditional probability of a sequence of labels given a sequence of observations, capturing complex dependencies between labels.

6.Deep Learning Models: Convolutional Neural Networks (CNNs) for spatial data (e.g., video frames) and Recurrent Neural Networks (RNNs) or Long Short-Term Memory (LSTM) networks for sequential data. Transformer networks are also being adapted for activity recognition due to their ability to capture long-range dependencies.

4. Algorithm Selection

The choice of algorithm depends on the specific requirements of the system:

- **Traditional Machine Learning**: Logistic Regression, Ridge Classifier, Random Forest, and Gradient Boosting are preferred for their simplicity, interpretability, and real-time performance. These algorithms are robust to noise, computationally efficient, and capable of handling a wide range of activity recognition tasks.
- **Deep Learning**: CNNs, RNNs, and Transformer networks provide superior accuracy for complex patterns but require large amounts of labeled data and substantial computational power. These are ideal for scenarios where high accuracy is critical, and resources are available.

V. Implementation

The implementation of the system leverages a robust combination of tools and techniques to ensure efficiency, accuracy, and scalability. Below, we detail the key components of the implementation:

1. Tools and Techniques

1. **Programming Language**: **Python** is the primary language due to its versatility, extensive libraries, and strong support for machine learning and data analysis. Its simplicity and readability make it ideal for implementing complex algorithms and integrating various system components.
2. **User Interface**: **Tkinter** is used to create an intuitive and user-friendly graphical interface. Its cross-platform compatibility ensures the system can be deployed across different operating systems without significant modifications.
3. **Modeling Framework**: **TensorFlow** serves as the backbone for developing and training machine learning models. Its comprehensive ecosystem provides powerful tools for building, optimizing, and deploying both deep learning and traditional machine learning models.
4. **Human Pose Estimation**: **MediaPipe** is used for real-time human pose estimation and body key point detection. Its pre-trained models enable accurate and efficient extraction of body landmarks, such as joints and skeletal structures, from video input.

2. Machine Learning Algorithms

The system incorporates a range of machine learning algorithms, each chosen for its complementary strengths:

- **Logistic Regression**: Offers simplicity and interpretability, making it ideal for baseline activity classification.
- **Ridge Classifier**: Introduces regularization to handle multicollinearity and prevent overfitting, enhancing model robustness.
- **Random Forest**: Excels at capturing complex, non-linear relationships in the data, making it suitable for intricate activity patterns.
- **Gradient Boosting**: Provides high predictive accuracy by sequentially improving model performance, particularly for sequential and fine-grained activity recognition.

3. Dataset Generation

To ensure the system's effectiveness and generalizability, a diverse and representative dataset is crucial. The dataset includes:

- **Normal activities** (e.g., walking, sitting, running).
- **Suspicious behaviors** (e.g., loitering, sudden movements).
- **Varied environmental conditions** (e.g., lighting, occlusion).

Activities are performed by multiple individuals under varying conditions to simulate real-world scenarios. This approach ensures the dataset captures a wide range of human movements and behaviors, accounting for differences in posture, speed, and interaction with the environment.

4. High-Level Enhancement

The integration of advanced tools, cutting-edge frameworks, and carefully curated datasets forms the foundation of a highly capable and adaptable system. Key enhancements include:

- **Seamless Development**: The use of Python, Tkinter, and TensorFlow ensures a smooth development process.
- **Balanced Approach**: The selection of machine learning algorithms provides a balanced approach to activity recognition and suspicious behavior detection.
- **Real-World Generalizability**: The emphasis on diverse sampling and real-world conditions enhances the system's ability to perform reliably across different settings.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 03 | March - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Output:

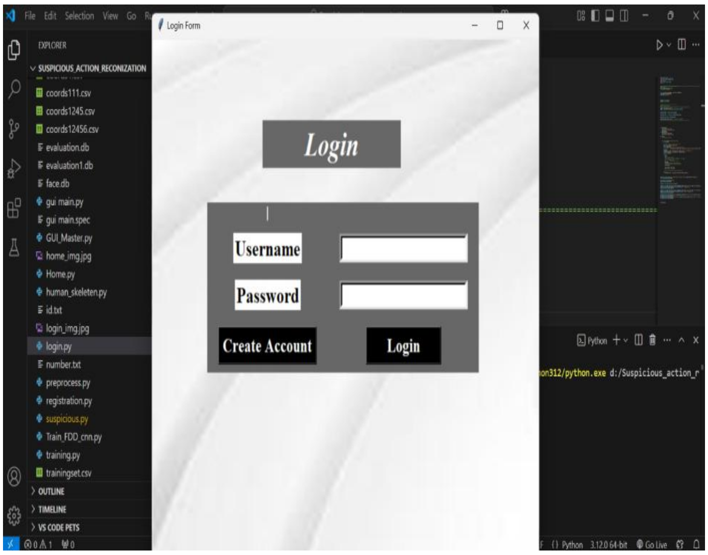

Fig.2: Login Page

Fig 3: Registration Page

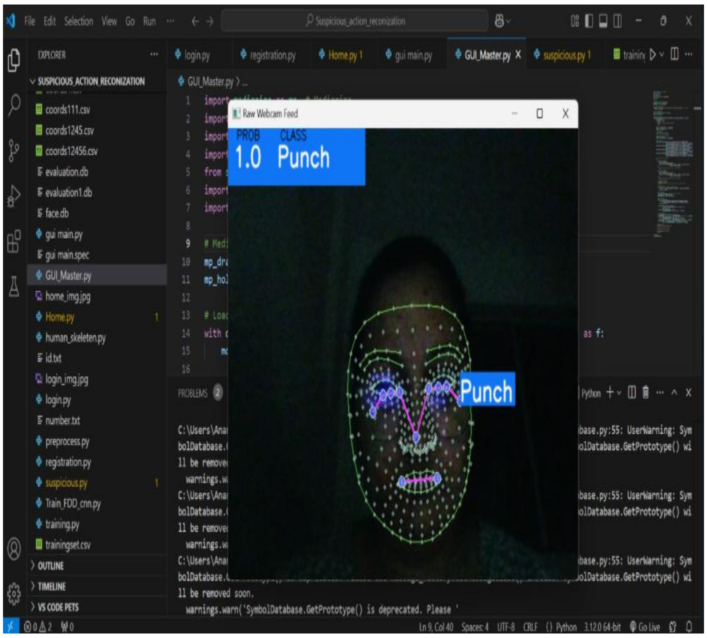

Fig 4:Live Activity Detection

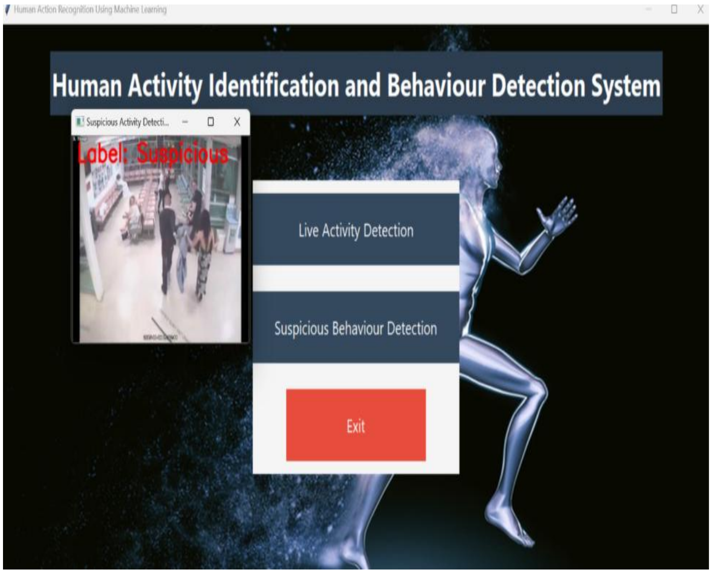

Fig 5: Suspicious Activity Detection

VI. CONCLUSION

This research presents an advanced system for real-time suspicious activity detection, initially designed for academic settings but with scalable potential for public and private spaces. By leveraging machine learning and computer vision, the system transcends the limitations of traditional CCTV surveillance, which often relies on post-event analysis. It proactively identifies suspicious behaviors and individuals, enabling timely interventions to prevent incidents before they escalate. This shift from reactive to proactive monitoring represents a significant advancement in security technology, offering a robust, adaptable, and efficient solution for enhancing safety across diverse environments. As the system evolves, its applications can expand, making it a vital tool for fostering safer communities and safeguarding assets in an increasingly complex world.

VII. REFERENCES

1.Contact Part Detection From 3D Human Motion Data Using Manually Labeled Contact Data and

Deep Learning, Authors: Changgu Kang, Meejin Kim, Kangsoo Kim, Sukwon Lee, Publisher: IEEE Year: 2023

2.Direction-Independent Human Activity Recognition Using a Distributed MIMO Radar System and

Deep Learning, Author: Sahil Waqar, Muhammad Muaaz, Matthias Pätzold,Publisher: IEEE,Year: 2023

3.Human-Based Interaction Analysis via Automated Key Point Detection and Neural Network Model

Authors: Israr Akhter, Naif Al Mudawi, Bayan Ibrahim AlAbdullah, Mohammed AlOnazi, Jeongmin Park, Publisher: IEEE, Year: 2023

4. Tripathi, Rajesh Kumar, Anand Singh Jalal, and Subhash Chand Agrawal. "Suspicious human activity recognition: a review." *Artificial Intelligence Review* 50 (2018): 283-339.

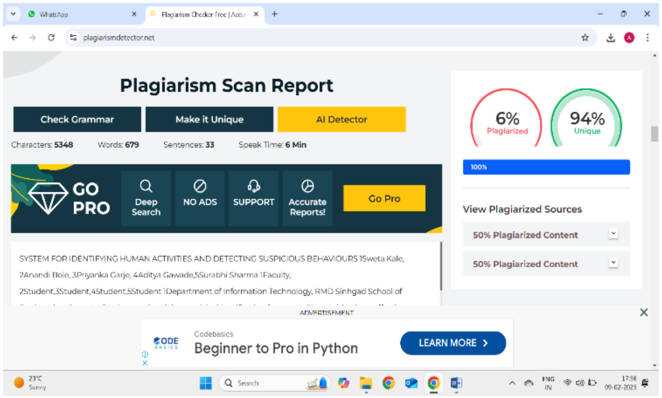

Fig 6: Plagiarism Check Result