2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

Real-Time Ragging Notification Using Machine Learning Algorithms and MediaPipe

R Dhanesh
PG Student
Department of Computer Applications
Hindustan Institute of Technology & Science
Chennai, India
22248012@student.hindustanuniv.ac.in

Dr. HJ Shanthi
Associate Professor
Department of Computer Applications
Hindustan Institute of Technology & Science
Chennai, India
hjshanthi@hindustanuniv.ac.in

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10627658

Abstract— **Ragging is disorderly practice of senior students harassing their juniors that violates rules of the institutions. Addressing critical issues such as ragging, suicidal ideation, and medical emergencies within college campuses necessitates the integration of innovative technological solutions. The utilization of innovative technologies such as MediaPipe and machine learning algorithms plays a pivotal role in addressing critical issues like ragging, suicidal ideation, and medical emergencies within college environments. This research paper integrates OpenCV, machine learning algorithms, MediaPipe, including support vector machine, random forest, k-nearest neighbour, and support vector machine, to develop a model for hand tracking and key point recognition. Through the deployment of a camera-enabled device, real-time footage is captured and compared against the trained model. Predictions and classification is detected from hand gestures, and when the sign matches the designated sign then it triggers an alarm and sends a screenshot of the person who needs help. This ensures prompt alerts authorities to take necessary actions in response to identified situations, thereby enhancing campus safety and welfare.**

Keywords—Ragging, MediaPipe, OpenCV, Support Vector Machine, Hand Sign Recognition, Random Forest, K-Nearest Neighbour, Stacking Classifier

I. INTRODUCTION

Machine Learning and Deep Learning technologies can be utilized to prevent and address ragging incidents in colleges. Monitoring systems such as surveillance systems with cameras and audio sensors can help detect unusual or aggressive behavior in common areas. This information can guide preventive measures. Implementing anonymous reporting systems where students can report incidents without fear of reprisal. ML algorithms can help analyze these reports to identify patterns or trends. ML models can analyze behavioral patterns among students to identify potential aggressors or victims. Unusual interactions or sudden changes in behaviour could be flagged for further investigation. Implementing a sign-based ragging detection system within a college environment requires a multi-faceted approach that integrates Machine Learning (ML) and Deep Learning (DL) technologies. Initiating this process involves collaboration with experts and professionals in psychology and counselling to define specific behavioural, verbal, and physical signs associated with ragging. These signs may encompass sudden changes in behavior, distress signals, verbal abuse, or physical aggression. Once signs are clearly defined, the data collection phase begins, wherein information on historical ragging incidents is meticulously gathered. This dataset serves as a foundation for training ML models, with a focus on extracting relevant features. Features may include patterns of communication, the frequency of certain behaviors, or specific locations where incidents tend to occur. The subsequent step involves training ML models, such as decision trees, support vector machines, or ensemble models, utilizing the extracted features and historical data. These models are designed to learn and recognize patterns associated with ragging signs. Integration of sensors, such as cameras and microphones, into common areas of the college campus facilitates real-time data collection. This data is then analyzed by the ML model for signs of ragging. To ensure timely intervention, an alert system is established. When the ML model identifies potential ragging signs during real-time monitoring, notifications are triggered and sent to designated authorities. Quick response is crucial in preventing the escalation of incidents. Continuous learning is implemented through a feedback loop, allowing the system to adapt and improve over time by incorporating new data and feedback from actual incidents. This research investigates the development of a real-time sign language recognition system specifically designed for the detection of ragging incidents. By employing advancements in machine learning and sign language recognition, the proposed system aims to achieve two primary objectives. Firstly, it seeks to enhance the accuracy of sign language recognition, leading to a more robust identification of potential ragging gestures. Secondly, it focuses on optimizing real-time processing capabilities to facilitate immediate response and intervention during such occurrences. The core of this system will be a meticulously designed model that integrates MediaPipe for hand tracking and leverages various machine learning algorithms for key point recognition within the signing gestures. This combined approach will ensure efficient capture and analysis of sign language for accurate detection. Ultimately, the system strives to provide prompt and reliable alerts to designated authorities, enabling timely intervention and a more effective response to ragging incidents. Implementing a sign-based ragging detection system within a college environment requires a multi-faceted approach that integrates Machine Learning (ML) and Deep Learning (DL) technologies. Initiating this process involves collaboration with experts and professionals in psychology and counselling to define specific behavioural verbal and physical signs associated with ragging. These signs may encompass sudden changes in behavior distress signals verbal abuse or physical aggression. Once signs are clearly defined the data collection phase begins wherein information on historical ragging incidents is meticulously gathered. This dataset serves as a foundation for training ML models with a focus on extracting relevant features. Features may include patterns of communication the frequency of certain behaviors or specific locations where incidents tend to occur. The subsequent step involves training ML models such as decision trees support vector machines or ensemble models utilizing the extracted features and historical data. These models are designed to learn and recognize patterns associated with ragging signs.

A. Artificial Neural Networks :

An Artificial Neural Network (ANN) is a computational model designed to mirror the intricacies of the human brain. Comprising interconnected nodes, or neurons, ANNs process information through layers—beginning with an input layer receiving external data. The data then flows through hidden layers where neurons conduct complex computations, leading to the final output layer that produces the network's result or prediction. This structure allows ANNs to excel at tasks such as pattern recognition, classification, and regression, making them versatile tools in various fields.

Z = Bias + W1X1 + W2X2 + …+ WnXn

- Bias: W0
- W: Weights
- X: Independent variables

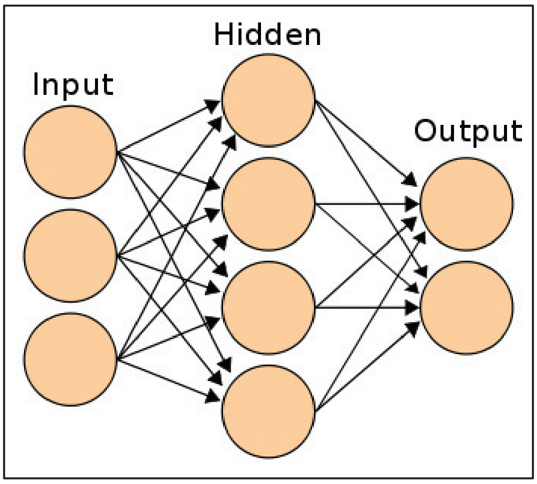

Fig. 1. Artificial Neural Network

B. Support Vector Machine:

This research proposes the adoption of Support Vector Machines (SVM) to classify hand gestures in sign language, leveraging their effectiveness in constructing optimal decision boundaries within high-dimensional feature spaces. Emphasizing both linear and non-linear classifications, the algorithm employs a Gaussian kernel for nuanced relationships within the data. Utilizing the Support Vector Classifier (SVC) from scikit-learn, key parameters such as the kernel type ('rbf'), regularization parameter (C=10), and gamma are specified. Success is contingent on meticulous parameter tuning and high-quality feature extraction, positioning this study as a contribution to advancing sign language interpretation through the robust capabilities of SVMs in hand gesture classification.

$X \cdot w - c \geqslant 0$

Putting c=b, we get:

$X \cdot w + b \geqslant 0$

Then, It results in:

Y = +1 , if $X \cdot w + b \geqslant 0$ and

Y = -1 , if $X \cdot w + b < 0$

In summary, this research aims to enhance sign language interpretation by employing SVMs for hand gesture classification, highlighting their efficacy in delineating decision boundaries and accommodating non-linear patterns. The study underscores the significance of parameter tuning and feature extraction for optimal model performance.

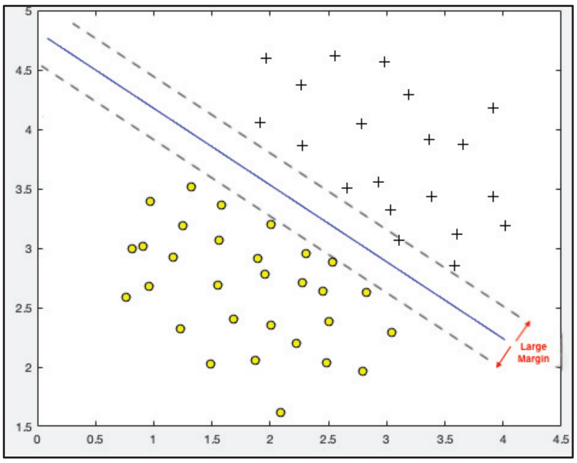

Fig. 2. Support Vector Machine

C. K-Nearest Neighbors:

In the realm of sign language recognition, the K-Nearest Neighbors (KNN) algorithm proves invaluable. KNN operates as a non-parametric, instance-based method, adept at classifying hand gestures into corresponding signs. Its core principle involves measuring the similarity between a test sample and training samples, subsequently selecting the K most similar instances for classification via a majority voting mechanism. While KNN demonstrates versatility, excelling in both regression and classification tasks, its preference leans towards the latter. This preference stems from its reliance on the assumption that similar data points congregate together. In practical applications like sign language recognition, KNN's simplicity and effectiveness make it a popular choice, despite sensitivity to parameters such as the number of neighbors and distance metrics.

Distance Functions:

1. Minkowski distance –

d(x,z) = (∑ r=1d|xr−zr|p)1/p

2. Manhattan distance –

d(x,y) = |x1 - x2| + |y1 - y2|

2. Euclidean distance –

d = √[(x2 2 − x1 1)2 + (y2 2 − y1 1)2]

Operating within the family of "lazy learning" models, KNN distinguishes itself by storing the entire training dataset in memory, avoiding a formal training stage. Despite its simplicity and accuracy, KNN encounters efficiency challenges as datasets grow larger, impacting overall model performance. Nevertheless, it remains a fundamental algorithm in data science, finding applications in recommendation systems, pattern recognition, financial market predictions, and various other domains. The initial ideas behind KNN, credited to Evelyn Fix and Joseph Hodges, contribute to its historical significance in the field.

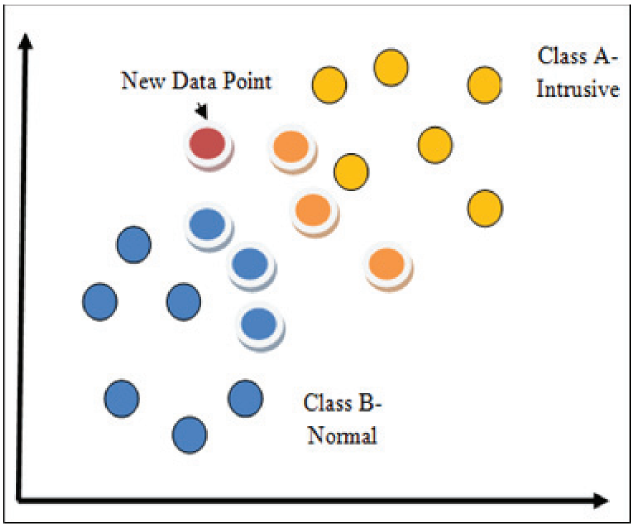

Fig. 3. K-Nearest Neighbors

D. Random Forest:

The research explores the application of the Random Forest algorithm in sign language recognition, emphasizing its suitability for handling intricate, non-linear relationships within high-dimensional datasets that incorporate multiple modalities. By constructing an ensemble of decision trees, the algorithm demonstrates its efficacy in mitigating overfitting and improving accuracy, a critical aspect when dealing with complex sign language patterns. The study specifically investigates the algorithm's performance on datasets comprising sign language videos or images, highlighting its ability to recognize diverse signs through the collaborative predictions of the decision tree ensemble.

II. RELATED WORKS

[1] The paper presents a new sign language dataset for motion generation, with 754 signs in 6786 videos. It addresses a MediaPipe detection issue for mirrored hands and reports a preliminary DTW value of 0.37 ± 0.23, acknowledging limitations and proposing future work to enhance motion generation using a robust transformer architecture. [2] The paper summarizes the ChaLearn LAP Large Scale Signer Independent Isolated SLR Challenge, involving 190 participants and 1.5K submissions. The paper suggests future research priorities such as large-scale datasets and continuous sign language, while the top-2 winning solutions, achieving over 96% recognition, leverage Graph Convolutional Networks (GCN) for effective spatiotemporal modeling. [3] The research examines ragging in Sri Lankan universities, revealing its complex nature as a power expression and response to social inequalities. The study, using theoretical frameworks, emphasizes the normalization of violence within the university subculture, with students trivializing its severe consequences. The cyclical nature of ragging, where victims become perpetrators, highlights the need for comprehensive interventions. The research advocates for a holistic approach, including students' perspectives, to address the societal problem effectively and create safer educational environments. [4]. This study focuses on Sign Language Recognition, offering a cost-effective approach using the "Finger Spelling, A" dataset and artificial intelligence-based techniques. Employing a two-layer image processing method, including whole-image processing and hand landmark extraction, the proposed multi-headed convolutional neural network (CNN) model achieves an impressive 98.981% test accuracy on 30% of the dataset. Mitigating overfitting concerns, the model incorporates data augmentation and dynamic learning rate reduction. The study foresees its contribution to the development of an efficient human–machine communication system for the deaf-mute community. [5]. This paper introduces a lightweight YOLO v3 and DarkNet-53-based model for gesture recognition, omitting the need for preprocessing and image enhancement. The proposed model attains high accuracy in complex environments and low-resolution picture modes. Evaluated on a labeled hand gesture dataset, it achieves outstanding metrics: accuracy (97.68%), precision (94.88%), recall (98.66%), and an F-1 score (96.70%). Comparative analysis with SSD and VGG16 shows superior performance, ranging between 82-85% accuracy. The trained model is adept at real-time detection of static hand images and dynamic gestures in videos, presenting a promising solution for communication assistance for individuals with disabilities. [6]. This paper tackles real-time gesture recognition challenges in computer vision, emphasizing the complexity of extracting hands from cluttered backgrounds. The proposed system, utilizing a bare hand and laptop webcam, ensures flexibility and speed without relying on a database. By combining Convex Hull and Convexity Defects, the system successfully recognizes sixteen hand gestures, achieving a notable recognition rate of 97.5%. The study suggests future enhancements, such as incorporating both hands for increased gesture variety. Experimental results highlight optimal recognition rates in clear backgrounds with medium lighting, underlining potential considerations for enhancing system accuracy in future implementations. [7]. This research aims to enhance real-time human-computer interaction by recognizing symbols drawn in front of a standard camera without the need for specialized sensors. Utilizing mathematical algorithms such as Accumulated Weight, CSRT Tracker, and OpenCV for detection and tracking, the model achieves precision in drawing tasks. The recognition phase employs deep learning, resulting in an impressive 98.56% accuracy in classifying English alphabet characters and numerals. The approach outperforms previous methods, offering a user-friendly HCI solution without the reliance on additional sensors.

III. EXISTING SYSTEM

The present research landscape in the realm of sign language and gesture recognition reflects a diverse amalgamation of methodologies and challenges. The existing process is shown in Fig. 3. Within this domain, researchers have introduced novel sign language datasets, aiming to augment motion generation and ameliorate detection challenges through the implementation of robust transformer architectures. However, it is paramount to underscore that these proposed enhancements are currently regarded as nascent developments, necessitating further refinement in subsequent investigations. In the context of large-scale sign language recognition competitions, prevalent systems have elucidated salient challenges and identified strategic priorities. Predominant among these solutions is the notable reliance on Graph Convolutional Networks (GCNs) for effective spatiotemporal modeling. Nevertheless, it is prudent to acknowledge that such seemingly efficacious solutions are not impervious to inherent limitations, thereby warranting ongoing refinement and development. In the field of gesture recognition, a meticulous survey of recent advancements in deep learning-based models, encompassing CNNs, RNNs, LSTMs, and Transformers, has been conducted. While diverse gesture recognition systems have been propounded, demonstrating commendable accuracy and efficiency in temporal processing, they too confront challenges. These challenges encapsulate the imperative for a substantive volume of data for effective training, the prospect of refining simultaneous gesture detection, and an ongoing exploration of methodologies designed to curtail training time. The current research panorama thus epitomizes a dynamic interplay between strides of progress and a cognizant acknowledgment of extant limitations, thereby impelling researchers toward the pursuit of innovative solutions and the resolution of both technical and societal challenges intrinsic to sign language and gesture recognition.

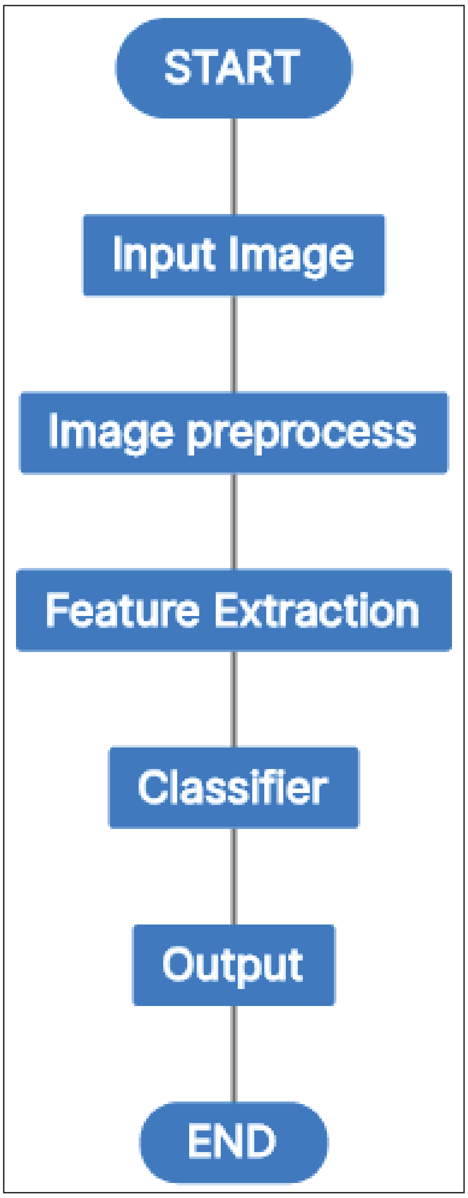

Fig. 4. Existing Architecture

IV. METHODOLOGY

The process to make alert the management of the institute or organization when a student needs help by recognizing the had sign made by the student are Data acquisition, Data preprocessing, Model Training and Gesture classification.

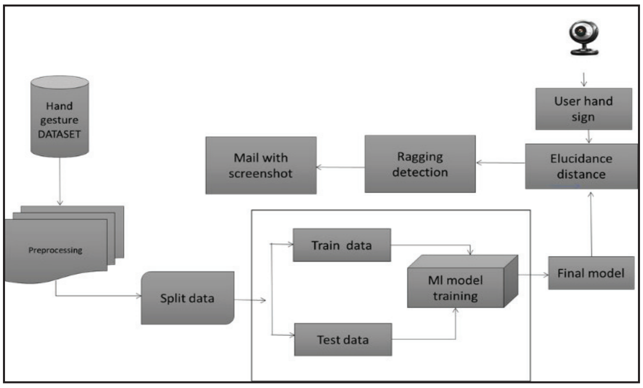

Fig. 5. Proposed Architecture

The architecture consists of Data collection, Preprocessing, Model training using Machine Learning Algorithms, Evaluating the metrics as shown in Fig. 4.1. The proposed approach involves using hand key points as input features for a machine learning model. Different algorithms, such as random forest, support vector machine, hybrid stacking model, and k-nearest neighbor, will be used. The model will be trained extensively, and its performance will be assessed using metrics like F1 score, precision, and recall. Selecting the best-performing model based on these metrics is vital. After model selection, the research aims to predict sign language gestures in real-time by comparing hand key points detected by MediaPipe with the trained model. This will include capturing video input from a webcam using OpenCV. Predicted signs may prompt authorities to take proactive measures via alert emails. This comprehensive approach covers both the technical aspects of model training and evaluation, as well as practical application in real-world scenarios to improve communication outcomes.

A. Data Acquisition:

The data collection process involves acquiring images of hand gestures, which serve as the primary dataset for the study. A dedicated Sign Language Recognition Module has been developed to discern and interpret the sign language gestures performed by users, translating them into numerical values. This transformation is executed by converting the images into numerical representations. The resulting dataset structure

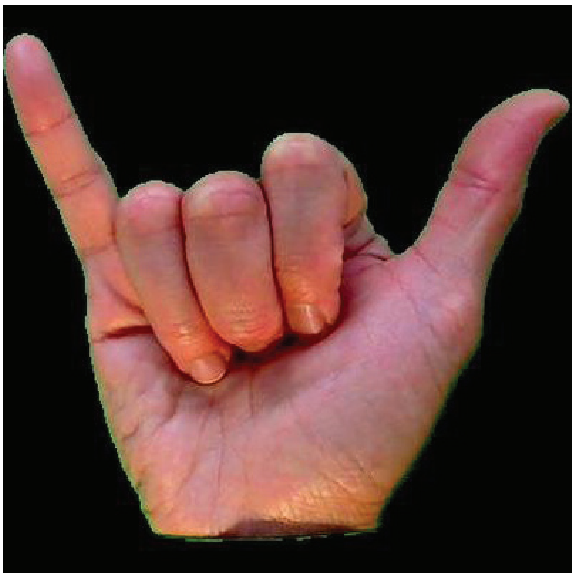

Fig. 6. Emergency Hand Sign

includes a column named "emergency_sign" shown in Fig. 4.2 denoting the type or label of the emergency sign. Additionally, distinct columns labeled "unit-0" through "unit-27" are incorporated to signify different units or features associated with each specific emergency sign. This meticulous approach to data collection ensures a comprehensive and standardized representation of the sign language gestures, facilitating subsequent analysis and recognition efforts.

B. Data Preprocessing:

The initial step in the data collection process involves a comprehensive examination of the hand gesture images ensuring a meticulous scrutiny of individual data points to guarantee their accuracy and completeness. Following this, a crucial aspect of the process is the identification of dependent and independent variables, where the former represents the predicted outcome, and the latter serves as predictors. This step is fundamental in structuring the dataset effectively for subsequent analyses. Subsequently, the dataset undergoes a systematic division into training and testing subsets. This strategic partitioning is pivotal for strengthening the model's robustness and its ability to generalize to newly generated and real-time data. The separation into training and testing data sets enables a thorough evaluation of the model's performance under different hand gestures, enhancing its overall reliability. Furthermore, a rigorous examination is conducted to identify and ad-dress any instances of missing values within the dataset. These systematic steps collectively contribute to establishing a solid foundation for subsequent phases of the research, particularly in the recognition of an accurate and reliable hand gesture recognition model.

C. Model Training:

The research delves into the application of machine learning algorithms, specifically Random Forest, K-Nearest Neighbors, Support Vector Machine, and a hybrid stacking model, for the identification of sign language gestures through hand gesture images. Random Forest, chosen for its noise-handling capability, constructs multiple decision trees, aggregating predictions for regression or classification tasks. SVM, a robust algorithm, identifies optimal boundaries in the feature space, accommodating both linear and non-linear decision boundaries. KNN, a simple yet effective approach, classifies based on the majority class of the K nearest neighbors in feature space, making it suitable for sign language recognition. The hybrid stacking model, leveraging the strengths of diverse models, enhances overall predictive accuracy through meta-model training. In practical implementation, the research employs code for the Random Forest Classifier, setting the test and train data and evaluating accuracy. Additionally, it discusses the real-time prediction aspect, integrating OpenCV to capture video input from a webcam and triggering proactive responses via alert emails based on the predicted sign language gestures. The comprehensive approach spans both technical aspects, such as model training and evaluation using metrics like F1 score, precision, and recall, and practical application, showcasing the potential for improved communication outcomes in real-time scenarios.

D. Model Testing:

The evaluation of machine learning models is a crucial step in assessing their performance. In the specific scenario of sign language recognition with Random Forest, K-Nearest Neighbors (KNN), and Support Vector Machine (SVM).

Conducting Model Evaluation helps to assess the models' performance using the designated testing set. Common metrics for evaluating classification tasks, such as accuracy, precision, recall, and F1 score, should be employed. Additionally, a thorough examination of the confusion matrix is essential to identify classes that may be frequently confused by the models during testing. This comprehensive evaluation provides insights into the effectiveness and potential limitations of the machine learning models in the context of sign language recognition. The evaluation metrics that will be considered are:

1) Confusion matrix:

The confusion matrix is a fundamental tool used in the evaluation of machine learning models, including those in the context of sign language recognition explored in the research paper.

2) Accuracy:

Accuracy in model training gauges the proportion of correctly predicted instances, serving as a fundamental metric for assessing the model's overall correctness.

$$\frac{TP + TN}{TP + FP + TN + FN}$$

3) Precision:

Precision quantifies the percentage of true positive instances relative to all instances predicted as positive by the model. It provides insight into the model's accuracy specifically when it claims positive identifications, highlighting its precision within the entire dataset.

$$\frac{TP}{TP + FP}$$

4) Recall:

Recall assesses the percentage of actual positive instances correctly identified by the model out of the total positive instances in the dataset. This metric reveals how many true positive instances the model missed when correctly identifying positive cases, providing insights into its ability to capture all relevant positives.

$$\frac{TP}{TP + FN}$$

5) F1 score:

The F1 score is a harmonic mean of precision and recall, considers the balance between the two metrics, emphasizing the need for both to be high for a better overall performance. The F1 score is particularly sensitive to decreases in either precision or recall, making it a comprehensive metric.

$$\frac{2}{\frac{1}{precision} + \frac{1}{recall}} = \frac{2 * precision * recall}{precision + recall}$$

E. Output prediction and Notification:

Once the hand gesture is recognized, then an email is sent to concerning authorities to promptly addressing instances of students in distress. The system incorporates advanced functionalities such as dynamic configuration of Gmail credentials, encompassing sender and receiver email addresses, along with secure authentication through application-specific or regular passwords. Leveraging MIME objects, the module efficiently constructs detailed email messages containing crucial information about the distressed student, utilizing the MIMEMultipart container for organization and MIMEText for the body text. Additionally, the system supports the seamless attachment of pertinent files or images, enhancing the comprehensiveness of the alert communication. The communication process is fortified through the establishment of a secure connection to the Gmail SMTP server, utilizing port 587 and initiating a TLS session to ensure encrypted and authenticated data transmission. The sender's login credentials, consisting of the Gmail email address and a designated application-specific password, are securely managed for authentication. The email dispatch mechanism, executed via the sendmail method, guarantees the confidentiality and integrity of the alert message during transmission. In conclusion, this research provides a technically advanced and se-cure framework for institution to deploy an email-based alert system, enabling proactive and effective communication to address instances of student distress and fostering a technologically sophisticated approach to student well-being management.

V. PROPOSED WORK

The proposed research utilizes a cutting-edge approach to enhance student safety in educational institutions through the integration of advanced technologies and machine learning. The experimental architecture involves key stages, including Data Collection, Preprocessing, Model Training, and Evaluation, employing innovative methods such as hand key points and diverse machine learning algorithms. In this chapter, a concise overview will be provided on the various scientific techniques and methodologies employed in this thesis. This discussion aims to enhance the reader's comprehension of both the implementation process and the resultant outcomes. The chapter will delve into fundamental theories in computer vision, machine learning, and MediaPipe, elucidating the specific computer vision techniques and methodologies applied in the execution of this project. The focus will be on advanced hand tracking and key point detection technologies, real-time gesture processing, and the integration of machine learning algorithms to address critical safety concerns on college campuses. By understanding these foundational concepts, readers will gain insights into the development and effectiveness of the proposed system for enhancing campus safety and well-being.

A. *Enhanced Sign Recognition Accuracy:*

This heading refers to the system's ability to accurately interpret sign language gestures made by individuals. By integrating advanced technologies like MediaPipe for hand tracking and key point detection, the system improves its accuracy in recognizing various aspects of hand motion, including shape, position, and trajectory. This enhancement ensures that the system can reliably identify specific signs or gestures despite the variability and complexity of hand movements.

B. *Real-Time Gesture Processing:*

The system's capability to process sign language gestures in real-time, providing timely feedback and facilitating smooth communication between users. By leveraging technologies such as MediaPipe and OpenCV, the system minimizes processing latency while maintaining acceptable levels of accuracy. This feature ensures that users receive immediate responses to their gestures, enhancing the overall user experience.

C. *Mitigation of Limited Annotated Data:*

The system utilizes various machine learning algorithms and rigorous testing methodologies to effectively mitigate this limitation. By selecting the optimal model and optimizing its performance using evaluation metrics, the system improves its ability to generalize to unseen gestures or sign variations, despite the limited availability of annotated data.

D. *Timely Incident Communication:*

This heading focuses on the system's capability to promptly communicate incidents such as ragging or emergencies to relevant authorities. By incorporating a database matching mechanism, the system can recognize predefined sign language gestures indicative of incidents and trigger alarms r send emails with screenshots to notify authorities. This feature ensures that incidents are promptly detected and reported, allowing for timely intervention and response, thus enhancing campus safety.

E. *Cost Effective Implementation:*

This heading emphasizes the system's cost-effectiveness in terms of both installation and maintenance. Unlike traditional sensor-based systems, which may require expensive equipment and infrastructure, the proposed system utilizes existing hardware such as webcams. By leveraging readily available technology and minimizing the need for additional sensors, the system offers a cost-effective solution for enhancing safety and well-being on college campuses, making it more feasible and sustainable in the long run.

VI. RESULT ANALYSIS

TABLE I. ACCURACY AND PRECISION METRICS

Algorithm / Metrics	Accuracy	Precision
Random Forest	98.41%	97.93%
KNN	94.45%	94.40%
SVM	96.83%	96.80%
Stacking Classifier	98.15%	98.06%

TABLE II. RECALL AND F1-SCORE METRICS

Algorithm / Metrics	Recall	F1-Score
Random Forest	97.46%	97.45%
KNN	95.06%	94.26%
SVM	97.06%	96.60%
Stacking Classifier	98.20%	98.00%

The Stacking Classifier model achieved a 98% accuracy in hand gesture recognition through its ensemble approach, leveraging decision trees for robust classification. Key to its success was training on a diverse dataset, enabling the model to generalize across various hand poses and backgrounds. Feature engineering and careful attribute selection enhanced the model's focus on essential information. Regularization techniques and hyperparameter tuning were employed to optimize performance and prevent overfitting. The result is a concise yet powerful model capable of accurately identifying gestures in real-world scenarios.

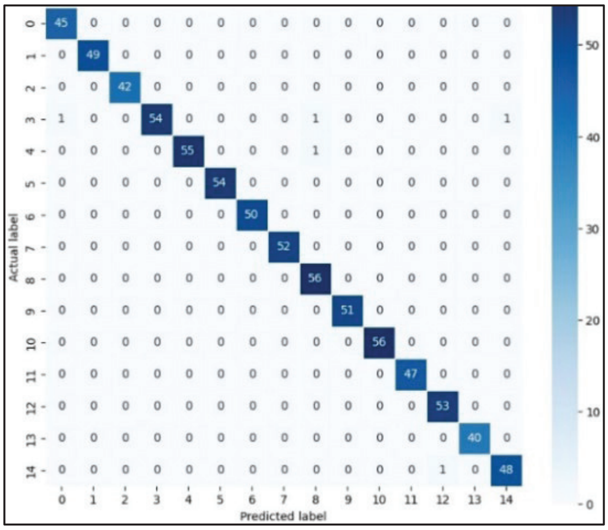

Fig. 7. Random Forest Confusion Matrix

A. *Random Forest Confusion Matrix:*

The confusion matrix (Fig. 6.1) for Random Forest provides a comprehensive overview of its performance in recognizing the specific hand sign displayed by students in ragging situations to alert authorities. It illustrates the algorithm's ability to accurately classify instances where the hand sign is present (true positives) and where it is absent (true negatives). Furthermore, it identifies any misclassifications, such as false positives and false negatives, offering insights for improvement. Random Forest demonstrates superior accuracy compared to other algorithms, such as KNN, in identifying both instances of the hand sign and non-occurrences.

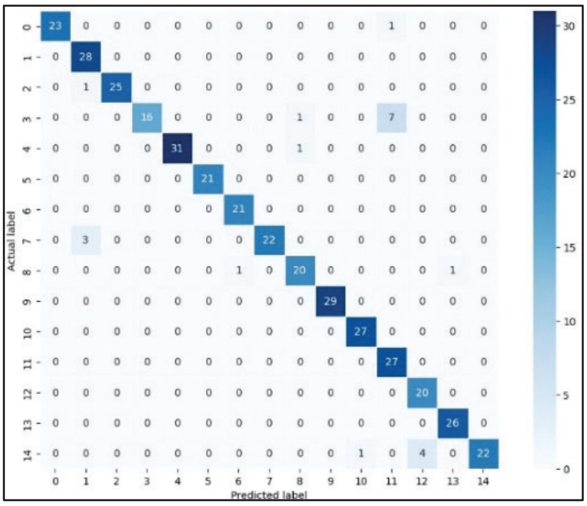

Fig. 8. KNN Confusion Matrix

B. *KNN Confusion Matrix:*

The confusion matrix (Fig. 7.2 KNN) for the K-Nearest Neighbors (KNN) algorithm provides a detailed breakdown of its classification performance. It offers insights into how effectively KNN distinguishes between ragging and non-ragging incidents, highlighting instances of correct classification as well as cases where misclassification occurs. Analyzing this matrix aids in assessing the algorithm's accuracy and identifying potential areas for optimization. While KNN demonstrates reasonable performance, it exhibits slightly lower accuracy compared to Random Forest, particularly in accurately identifying instances of ragging incidents.

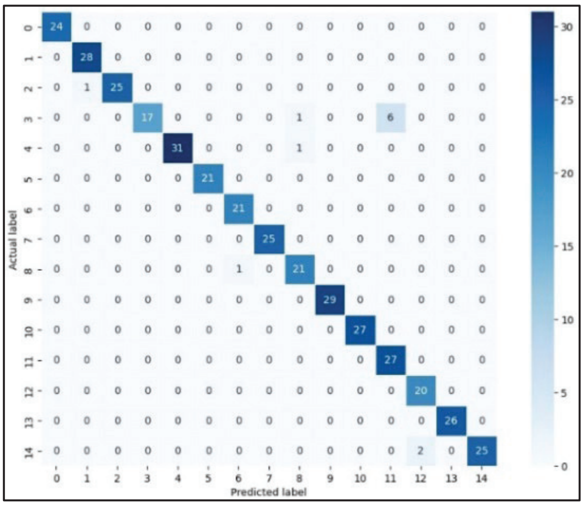

Fig. 9. SVM Confusion Matrix

C. *SVM Confusion Matrix:*

The confusion matrix (Fig. 6.3) for the Support Vector Machine (SVM) algorithm offers valuable insights into its classification performance within the ragging notification system. It showcases the algorithm's proficiency in accurately identifying instances of ragging incidents, demonstrating high counts of true positives and true negatives. Additionally, it highlights any instances of misclassification, providing a comprehensive evaluation of SVM's effectiveness in distinguishing between different types of incidents. Compared to KNN, SVM shows a higher level of accuracy in identifying ragging incidents, with fewer instances of misclassification.

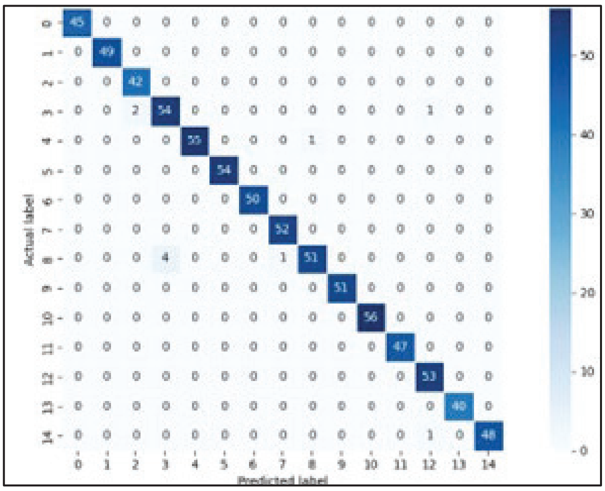

Fig. 10. Stacking Classifier

D. *Stacking Classifier Confusion Matrix:*

Finally, the confusion matrix for the Stacking Classifier presents a comprehensive assessment of its classification performance. It illustrates how effectively the Stacking Classifier combines multiple models to improve classification accuracy, showcasing instances of correct classification and highlighting any instances of misclassification. By analyzing this matrix, we can gain insights into the ensemble approach's effectiveness in enhancing the accuracy of the ragging notification system. The Stacking Classifier demonstrates superior performance compared to individual algorithms like KNN and SVM, leveraging ensemble learning techniques to achieve higher accuracy in identifying both ragging incidents and non-ragging situations.

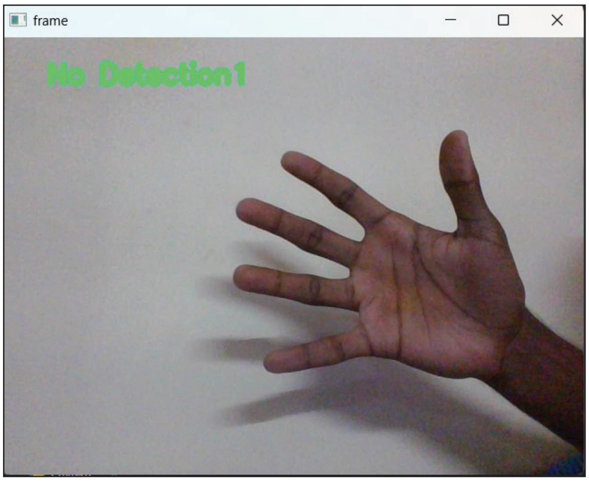

Fig. 11. Result when other hand signs are detected

E. When no Emergency sign is detected:

The Fig. 6.5 shows the response of the system when the emergency hand sign is not detected and it displays “No detection” on the screen. It checks for the sign continuously every second.

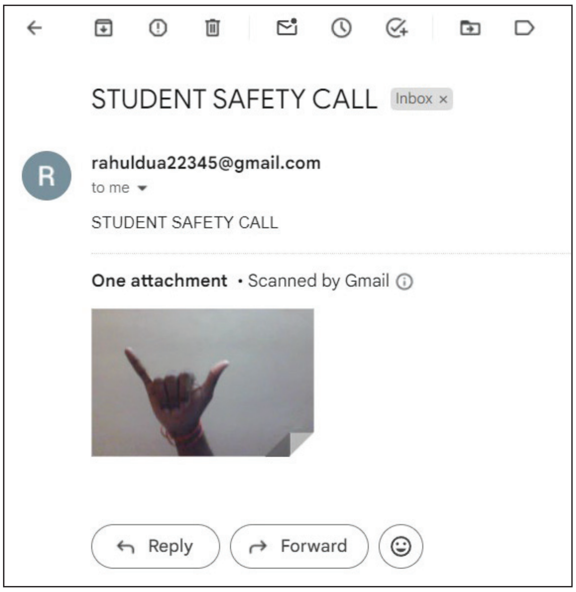

Fig. 12. Result when the “Emergency Hand Sign” is detected

F. When Emergency sign is detected:

The Fig. 6.6 shows the response of the system when the emergency hand sign is shown a detected, where a screenshot is taken from the camera at the very moment when the detection is made and the screenshot is sent to the desired authority as an attachment with an email.

VII. CONCLUSION

This research successfully integrates machine learning algorithms and MediaPipe technology to develop a real-time notification system for detecting ragging incidents on educational campuses. By employing different machine learning algorithms for hand gesture recognition, the system achieves high accuracy in identifying distress signals and promptly alerts authorities. The deployment of a camera-enabled device for real-time footage analysis, coupled with the trained model, ensures accurate prediction and classification of gestures. This immediate response capability is crucial for addressing critical issues such as ragging, suicidal ideation, and medical emergencies, thereby enhancing the safety and welfare of the campus community. The study demonstrates the significant role that advanced technological solutions can play in creating safer educational environments. The robust methodology and effective implementation underscore the potential of machine learning in proactively addressing and mitigating incidents that threaten student well-being.

VIII. FUTURE SCOPE

A. Diverse Detection Methods:

To overcome limitations in detecting ragging incidents, future enhancements can broaden the system's capabilities by incorporating a wider range of culturally relevant hand signals or alternative methods like hand raises. Additionally, context-awareness can be achieved by integrating signs with pointing gestures towards specific locations, providing more nuanced emergency information.

B. Multi-modal Integration:

Integration of facial recognition and audio analysis alongside hand gestures can significantly enhance accuracy and reduce false positives. Seamless integration with existing security infrastructure, such as automatic camera triggering and door locks based on emergency types, can bolster response mechanisms and provide a comprehensive safety ecosystem.

C. Wearable Technology Integration:

Future iterations may explore integration with wearable technology for discreet emergency alerts and vital sign monitoring, offering deeper insights into students' emotional states. This can contribute to proactive measures and timely interventions to ensure a secure learning environment.

D. Continuous Improvement:

Continuous model learning through real-world deployment data and user feedback will be crucial in refining the system's effectiveness over time. Additionally, complementary awareness campaigns and bystander intervention training can foster a culture of safety and support, further strengthening the safety ecosystem in educational institutions.

REFERENCES

[1] Cardoso Duarte, A., Palaskar, S., Ventura Ripol, L., Ghadiyaram, D., DeHaan, K., Metze, F., ... & Giró Nieto, X. (2021). How2Sign: A large-scale multimodal dataset for continuous American sign language. In 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition: Virtual, 19-25 June 2021: proceedings (pp. 2734-2743). Institute of Electrical and Electronics Engineers (IEEE).

[2] Sincan, O. M., Junior, J., Jacques, C. S., Escalera, S., & Keles, H. Y. (2021). Chalearn LAP large scale signer independent isolated sign language recognition challenge: Design, results and future research. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 3472-3481).

[3] Wickramasinghe, A., Axemo, P., Essén, B., & Trenholm, J. (2022). Ragging as an expression of power in a deeply divided society; a qualitative study on students perceptions on the phenomenon of ragging at a Sri Lankan university. PloS one, 17(7), e0271087.

[4] Pathan, R.K., Biswas, M., Yasmin, S. et al. Sign language recognition using the fusion of image and hand landmarks through multi-headed convolutional neural network. Sci Rep 13, 16975 (2023). https://doi.org/10.1038/s41598-023-43852-x

[5] Mujahid, A., Awan, M. J., Yasin, A., Mohammed, M. A., Damaševičius, R., Maskeliūnas, R., & Abdulkareem, K. H. (2021). Real-time hand gesture recognition based on deep learning YOLOv3 model. Applied Sciences, 11(9), 4164.

[6] AlSaedi, A. K. H., & AlAsadi, A. H. H. (2020, February). An efficient hand gestures recognition system. In IOP Conference Series: Materials Science and Engineering (Vol. 745, No. 1, p. 012045). IOP Publishing.

[7] Niloy, E., Meghna, J., & Shahriar, M. (2021, July). Hand gesture-based character recognition using opencv and deep learning. In 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI) (pp. 1-5). IEEE.

[8] Oudah, M., Al-Naji, A., & Chahl, J. (2020). Hand gesture recognition based on computer vision: a review of techniques. journal of Imaging, 6(8), 73.

[9] Nishani, E., & Çiço, B. (2017, June). Computer vision approaches based on deep learning and neural networks: Deep neural networks for video analysis of human pose estimation. In 2017 6th Mediterranean Conference on Embedded Computing (MECO) (pp. 1-4). IEEE.

[10] Sainadh, K. V., Satwik, K., Ashrith, V., & Niranjan, D. K. (2023, April). A Real-Time Human Computer Interaction Using Hand Gestures in OpenCV. In International Conference on Information and Communication Technology for Intelligent Systems (pp. 271-282). Singapore: Springer Nature Singapore.

[11] Kumar, N., Dalal, H., Ojha, A., Verma, A., & Kaur, M. (2023, November). Real-Time Hand Gesture Recognition for Device Control: An OpenCV-Based Approach to Shape-Based Element Identification and Interaction. In 2023 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS) (pp. 1537-1541). IEEE.