

Safe Alert: Enhancing Women's Security with ML and Gesture Recognition

Deekshmitha Reddy, Lavanya S, and Gowri Vidhya N
St. Joseph's institute of technology

Abstract—In response to the growing need for enhanced safety solutions for women, particularly in public spaces and at night, we present an innovative system leveraging advanced technologies such as MediaPipe, OpenCV, and machine learning algorithms. Our solution captures real-time video through a webcam and recognizes predefined sign language gestures to detect potential threats. Through an integrated database mechanism, the system triggers alarms and sends email notifications when matched gestures indicate danger, ensuring timely responses. We evaluate various machine learning models—Decision Tree, Gaussian Naive Bayes, and Gradient Boosting—based on performance metrics such as F1 score, precision, and recall to determine the most effective solution. The real-time gesture recognition and high-precision detection offer a valuable tool in improving women's safety in public spaces, while MediaPipe's hand-tracking capabilities enhance the accuracy of sign language recognition.

I. INTRODUCTION

In the modern world, rapid advancements in technology have not only transformed how we live but have also created new opportunities to address persistent social challenges. One such challenge, which remains a pressing issue globally, is women's safety. Women, especially those who navigate public spaces during late hours, face heightened risks of harassment, assault, and other forms of violence. While traditional safety measures like surveillance cameras and emergency helplines offer some degree of protection, they are often reactive in nature and may not provide timely intervention.

Our project aims to address this critical need by developing a real-time, technology-driven safety system that enhances the security of women in public spaces. By leveraging advanced tools such as MediaPipe, OpenCV, and various machine learning algorithms, our system is designed to proactively detect signs of distress and initiate immediate responses. The core functionality of the system is based on recognizing predefined sign language gestures that indicate danger, which are captured through a webcam and processed using real-time video analysis.

The system's recognition process utilizes hand-tracking capabilities provided by MediaPipe, along with keypoint detection, to accurately interpret sign language gestures. Once a gesture is recognized and matched against a predefined set of danger signals, the system triggers an alarm and sends an email notification to relevant authorities or emergency contacts. To ensure the system's responsiveness, a database matching mechanism is implemented, allowing for instantaneous alerts and action.

The project also explores the application of various models in machine learning—such as Decision Tree, Gaussian Naive Bayes, and Gradient Boosting—to enhance the accuracy of gesture recognition. The models are evaluated based on key performance metrics, including precision, recall, and finally F1 score, to conclude the most effective approach for the real-world implementation. Extensive testing and validation ensure that the system can operate reliably in diverse environments and under various conditions.

By integrating these cutting-edge technologies, our project not only offers a practical solution to improve women's safety but also raises awareness about the role of machine learning and real-time video analysis in addressing social issues. We believe that this system, with its ability to respond proactively to threats, can make a meaningful contribution to creating safer environments for women worldwide.

II. LITERATURE SURVEY

Hand Gesture Recognition (HGR) has garnered significant attention across various research domains, including human-computer interaction, sign language interpretation, and security applications. The methods used in HGR vary widely, from simple, rule-based systems to advanced machine learning algorithms. This review focuses on recent advancements in gesture recognition using various techniques, such as **Elliptical Fourier Descriptors**, **Convolutional Neural Networks (CNN)**, **Support Vector Machines (SVM)**, **Finite State Machines (FSM)**, and **Fuzzy Logic**.

1. Fourier Descriptors for Sign Language Recognition

CP.V.V. Kishore et al. [1] explored the application of **Elliptical Fourier Descriptors (EFD)** and **Artificial Neural Networks (ANN)** in a four-camera model for **sign language recognition (SLR)**. This approach aims to mitigate occlusion problems caused by one hand blocking the other during sign gestures. EFDs are effective in reducing noise and providing robust feature vectors for classification. The study highlights the key advantage of robustness to occlusions, but it also acknowledges that such techniques suffer from poor segmentation and erroneous classification under challenging conditions. Furthermore, ANN models require a significant amount of computational resources to train and optimize, posing another challenge for real-time systems.

2. CNN for Feature Extraction and Classification

Barbhuiya et al. [2] employed **Convolutional Neural Networks (CNN)** for **sign language recognition** focusing on

both characters such as alphabets and numerals of **American Sign Language (ASL)**. The authors used **AlexNet** and **VGG16** architectures to extract features followed by classification using a **Support Vector Machine (SVM)**. An approach based on CNN automatically detects relevant features without manual intervention, leading to high recognition accuracy. However, CNNs are computationally expensive and require large datasets for effective training. Moreover, CNNs lack spatial invariance, which can be a limitation in dynamic environments, such as gesture recognition in real-world conditions. Despite these drawbacks, CNNs have shown promise in image classification tasks, including HGR.

3. Weighted Support Vector Machines for Pattern Recognition

Xanthopoulos and Razzaghi [3] proposed a **Weighted Support Vector Machine (WSVM)** used for fault detection and **control chart pattern recognition**. They demonstrated that WSVM improves the classification accuracy in multi-class environments by handling class imbalance better than traditional SVM models. This technique performs well in managing with multidimensional and spontaneous features, making it applicable to various domains, including gesture recognition. However, like other SVM-based methods, WSVM requires a huge amount of sample size to accomplish optimal results and often faces challenges in tuning hyperparameters, limiting its applicability in resource-constrained environments.

4. Multilingual Sign Language Recognition

Tornay et al. [4] developed a **multilingual approach** for sign language recognition, leveraging **Hidden Markov Models (HMMs)** to model hand movements across various sign languages, including **Swiss German, German, and Turkish**. The HMM approach helps in eliminating label bias by using independent hand movement data. This multilingual approach performed well across different sign languages, providing a flexible framework for recognizing multiple sign languages with minimal adjustments. However, HMMs require a large amount of training data to achieve high accuracy, and the models often suffer from unstructured parameters that complicate their optimization.

5. Finite State Machines and Fuzzy Logic for Gesture Recognition

Verma and Dev [5] presented a **gesture recognition system** using **Finite State Machines (FSM)** and **Fuzzy Logic**. Their approach does not require precise inputs, making it ideal for real-time systems where hand gestures may vary in speed and form. **Fuzzy c-means clustering** is employed to categorize hand postures, which are then used to define states in FSMs. This combination of FSM and fuzzy logic allows for flexibility and robustness in the system, handling noisy and imprecise

data effectively. However, FSM-based systems are highly dependent on human expertise, and fuzzy logic models are often less accurate compared to machine learning algorithms. This limits their use in complex environments where high precision is required.

6. CNNs for Hand Gesture Recognition

Barbhuiya et al. [2] explored the use of CNNs for **Hand Gesture Recognition (HGR)** by employing modified **AlexNet** and **VGG16** architectures for feature extraction. Their research demonstrated the effectiveness of CNNs in handling image classification tasks, with high accuracy in both alphabet and numeral recognition for ASL. Despite the computational overhead and high resource demands, CNNs offer automatic feature extraction, making them highly suited for complex gesture recognition systems. However, the main limitations include the requirement for extensive training data and the inability to encode object orientation and spatial relations.

7. Meta-Analytics for Gesture Recognition

Simske [1] introduced **Meta-Analytics**, a novel approach that combines multiple analytics techniques to derive meaningful insights from data. This technique focuses on building hybrid systems that enhance the overall prediction performance by reducing variance in predictive errors. Meta-Analytics offers high accuracy and stability compared to traditional models, as it builds classifiers that integrate various statistical methods. However, one drawback is the increased complexity in model interpretation and the potential for overfitting or underfitting when ensemble learning techniques are not properly calibrated. The benefits of Meta-Analytics lie in its ability to provide more consistent results across diverse data sources, but careful attention must be paid to model validation and optimization. [10] presented an innovative visual aid framework for completely blind people, which takes the form of a pair of glasses. The following are some of the most essential characteristics of the proposed device. The complicated algorithm processing is carried out on the Raspberry Pi 3 Model B+, which has low-end computing power. Using a combination of camera and ultrasound sensors and GPS-based location tracking for use in a navigation system, this Internet of things-based device offers advanced dual detection and distance measurement capabilities. This device makes it possible to have better access, solace, and navigational ease to blind people

III. METHODOLOGY

This project focuses on enhancing **women's safety** by developing a **real-time incident notification system** based on **Sign Language Recognition (SLR)**. The system will detect hand gestures related to distress or behavioral issues using **hand key points** as input features for a machine learning model. The real-time analysis will be achieved using tools

such as **MediaPipe**, **OpenCV**, and **scikit-learn**, and the system will trigger alerts to authorities via email upon recognition of specific gestures.

1. Overview of the System Architecture

The proposed system consists of two key modules:

1. **Gesture Recognition System**
2. **Real-time Alert Notification System**

1.1. Gesture Recognition System

- **Video Capture:** Using OpenCV, the system captures real-time video input from a CCTV camera or webcam.
- **Hand Key Point Detection:** MediaPipe will be utilized to detect and extract hand key points from the video stream in real-time.
- **Feature Extraction:** The detected hand key points are used as features in input for the machine learning models to classify the recognized gesture.
- **Machine Learning Models:** Several machine learning algorithms, including **Decision Tree**, **Gaussian Naive Bayes**, **Gradient Boosting**, and a **Hybrid Stacking Classifier**, are trained using the extracted hand key point data to classify the gestures.
- **Gesture Prediction:** The selected machine learning model will be used to classify the gestures in real-time.

1.2. Real-time Alert Notification System

- **Behavioral Gesture Identification:** Specific gestures indicating distress or danger, which may correlate with behavioral issues, will be recognized by the trained SLR model.
- **Alert Trigger:** Upon recognizing such gestures, the system will send an **email alert** to authorities or other designated personnel. The alert contains the incident's video frame and the gesture identified.
- **Silent Alert Feature:** This "Silent Alert" functionality can be used in sensitive situations where the person in distress cannot communicate verbally but can use pre-defined gestures to signal for help.

2. Hand Key Point Detection Using MediaPipe

- **MediaPipe** is a powerful open-source library that provides highly efficient real-time tracking of human body landmarks, including hand key points. The following steps will be followed:
 - **Hand Landmark Detection:** Each frame from the video feed is processed to detect **21 hand landmarks** (key points) per hand. These landmarks include the positions of the fingers, palm, and wrist.
 - **Feature Vector:** The hand key points are represented as 3D coordinates (x, y, z).

These points serve as input features for training the machine learning model.

3. Machine Learning Models for Gesture Classification

Several models in machine learning will be trained and finally evaluated to classify the hand gestures. The methodology for training and evaluating these models is as follows:

3.1. Dataset Preparation

- **Data Collection:** A labeled dataset of various gestures, including those related to behavioral issues, will be collected. The data will consist of the hand key points detected by MediaPipe.
- **Preprocessing:** The hand key points will be normalized to ensure consistency across varying hand positions and orientations. Data augmentation techniques may be applied to improve model robustness.

3.2. Model Training

- **Decision Tree:** A simple decision tree classifier will be trained using the hand key points. It is quite easy to understand and can handle non-linear relationships but may overfit the data given in the training.
- **Gaussian Naive Bayes:** This probabilistic model will be applied to classify the gestures based on the likelihood of feature distributions. Naive Bayes is computationally efficient and performs well on small datasets.
- **Gradient Boosting:** A boosting algorithm that builds an ensemble of weak learners (decision trees). Gradient Boosting can handle complex gesture patterns and generally achieves high accuracy.
- **Hybrid Stacking Classifier:** A hybrid model combining the strengths of multiple algorithms. In this case, we will use **Decision Tree**, **Naive Bayes**, and **Gradient Boosting** serves as base learners, and a **Logistic Regression** serves as a meta-learner for stacking the predictions. This approach binds the different algorithm's strength to improve overall performance.

3.3. Model Evaluation

- **Cross-validation:** Each model will be evaluated using k-fold cross-validation to ensure generalizability.
- **Evaluation Metrics:** The models will be evaluated based on the following metrics:
 - **F1 Score:** Measures the balance between precision and recall and finally providing a single metric for model performance.
 - **Precision:** Measures the proportion of perfectly predicted positive instances among all positive predictions.

- **Recall:** Measures the proportion of perfectly predicted positive instances among all actual positive instances.

4. Model Selection and Optimization

Based on the performance of the models on the evaluation metrics, the best-performing model will be selected. The **Hybrid Stacking Classifier** is expected to outperform individual models due to its ability to leverage the strengths of multiple classifiers. The selected model will be optimized using hyperparameter tuning techniques such as **Grid Search**.

5. Real-time Prediction and Notification System

Once the best model is selected, it will be integrated into the real-time system. The methodology for real-time prediction and notification is as follows:

5.1. Video Stream Processing

- **OpenCV** will capture the video feed in real-time from the CCTV camera or webcam.
- The frames will be passed to **MediaPipe** for detecting hand key points.

5.2. Gesture Prediction

- The detected hand key points in each frame will be fed into the trained model to predict the corresponding sign language gesture.
- Upon recognizing gestures related to behavioral distress, the system will trigger the **Silent Alert** mechanism.

5.3. Alert Mechanism

- **Email Notification:** The system will use an email API (such as **SMTP**) to send an alert email to designated authorities or emergency contacts. The email will include details of the recognized gesture and a snapshot of the video frame where the gesture was detected.
- **Silent Alert:** This feature will ensure that alerts are sent discreetly, allowing individuals in distress to seek help without drawing attention to themselves.

6. Tools and Technologies

- **MediaPipe:** For hand key point detection and tracking.
- **OpenCV:** For capturing video feeds from CCTV or webcam.
- **scikit-learn:** For implementing and training machine learning models.
- **Python Libraries:** Including **NumPy** and **pandas** for data processing and feature extraction, **Matplotlib** for visualizing results, and **SMTP** for sending email alerts.

7. System Testing and Validation

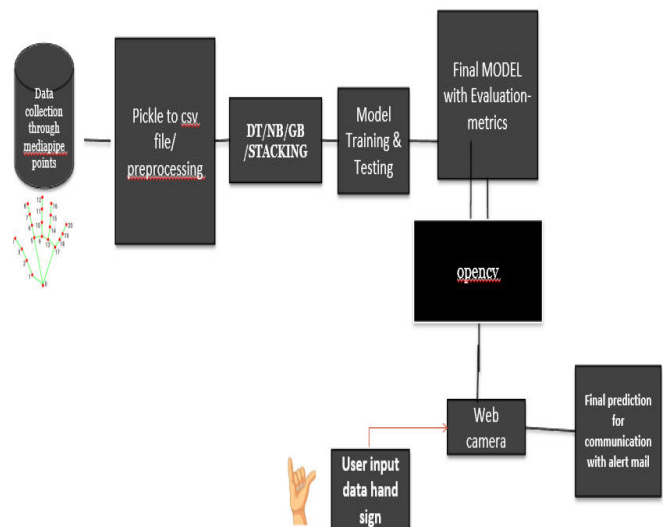
The system will be tested in different environments to validate its performance. The following scenarios will be considered:

- **Controlled Indoor Environment:** To ensure the system works reliably in settings such as homes, offices, or public spaces.
- **Outdoor Environment:** Testing the system's robustness to noise, varying lighting conditions, and background distractions.
- **Gesture Variability:** Evaluating the system's ability to handle variations in hand position, orientation, and speed during gesture performance.

8. Innovative Application: Silent Alert System

The **Silent Alert** functionality extends beyond conventional SLR applications by incorporating **AI-enhanced hand sign recognition** for real-time **incident notification**. This innovative approach addresses the growing need for non-verbal communication solutions in situations where verbal distress signals may not be feasible, particularly in cases involving women's safety. [8] discussed that K-means transformation, histogram equalization, linear contrast stretching, and share-based features are all used to detect leukemia. A method for automatically classifying leukocytes using microscopic images is proposed. This proposed model used MATLAB to find leukemia cells in healthy blood cells, and it requires no medical equipment or expert and heavily relies on automation. This technology can detect anemia, malaria, vitamin B12 deficiency, and brain tumors. The proposed method correctly identifies WBCs and leukoblasts in images and refines the identification, thresholding, and segmentation phases.

IV. SYSTEM ARCHITECTURE



V. RESULTS AND ACCURACY

Performance-Model Evaluation

To evaluate the performance of the machine learning models, we employed several metrics: Precision, Recall, and F1-Score. These metrics provide a complete assessment of each model's ability to precisely classify gestures associated with behavioral issues for women's safety.

1. **Precision:** It measures the accuracy of the positive predictions made by the model. It is calculated as the ratio of true positives (correctly identified gestures) to the sum of true positives and false positives (gestures incorrectly identified as positive). A higher precision indicates lesser false positives and more reliable predictions.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

2. **Recall:** Recall provides the model's ability to identify all of the relevant instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives (gestures that were particularly missed by the model). A higher recall means that the model is successfully detecting almost all of the relevant gestures.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

3. **F1-Score:** The F1-Score is generally the harmonic mean of Precision and Recall. It balances the trade-off between precision and recall, providing a single metric that reflects both aspects. This score is generally useful when dealing with imbalanced datasets.

$$\text{F1-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

4. **Accuracy:** Accuracy is calculated as the ratio of the number of correct predictions (both true positives and true negatives) to the total number of data instances. It provides an overall measure of the model's performance.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Data Instances}}$$

Model Comparison

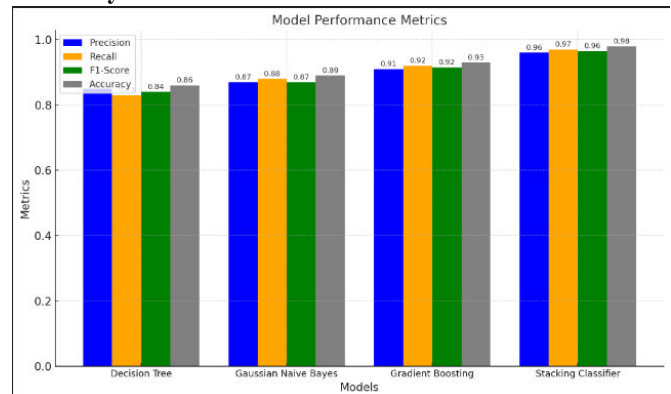
The performance of various algorithms, including Decision Tree, Gaussian Naive Bayes, Gradient Boosting, and a hybrid Stacking Classifier, was compared. Each model's precision, recall, F1-score, and accuracy were assessed to determine the best-performing approach. The Stacking Classifier, which combines multiple models to improve prediction accuracy, showed superior performance in handling the complex task of sign language gesture recognition in real time.

Innovations and Application

The selected model demonstrated high accuracy in real-time gesture recognition, enabling effective incident notification through the "Silent Alert" system. By processing video feeds from CCTV cameras with the SLR model, specific gestures associated with behavioral issues were identified, triggering alerts for immediate action. The integration of MediaPipe, OpenCV, and the trained model facilitated real-time monitoring and response, significantly enhancing women's safety.

Performance Metrics

The graph above illustrates the performance metrics of various machine learning models—**Decision Tree**, **Gaussian Naive Bayes**, **Gradient Boosting**, and **Stacking Classifier**—based on **Precision**, **Recall**, **F1-Score**, and **Accuracy**.



VI. DISCUSSION

Existing System:

1. Image Analysis:

- **Convolutional Layers (CNN):** Utilized for image recognition tasks, focusing on spatial features and patterns within images. CNNs are greatly effective in tasks such as detection of an object and classification of an image.
- **Inception Models (Inception-v3):** Enhance CNNs with multiple filter sizes

- and improved architectures for more accurate and efficient image classification.
- **RNN (Recurrent Neural Networks):** Proposed in a way for processing sequential data with temporal dependencies, making them suitable for tasks such as sequence prediction.
- **LSTM (Long Short-Term Memory):** An RNN variant that addresses long-term dependencies, commonly used in speech recognition and other sequential tasks.
- **Hybrid Architecture:** Combines CNN and RNN for video analysis, leveraging CNN for spatial feature extraction and RNN for temporal sequence processing.

Proposed System:

1. **Hand Key Points as Features:**
 - Uses hand key points detected by MediaPipe as input features for gesture recognition, focusing on spatial and positional data of hand gestures.
2. **Model Training and Evaluation:**
 - **Algorithms:** Includes Decision Tree, Gaussian Naive Bayes, Gradient Boosting, and a hybrid Stacking Classifier. These algorithms are trained on hand key point data to classify gestures.
 - **Evaluation Metrics:** Several metrics were used such as precision, recall, F1-score, and accuracy to assess model performance.
3. **Real-Time Gesture Recognition:**
 - **Video Input:** Uses OpenCV to capture webcam video and predict gestures in real-time.
 - **Alerts:** Integrates with a notification system to send alerts to authorities based on detected gestures.

Advantages of the Existing System

1. **Specialized Models:**
 - **CNNs:** Well-suited for high-performance in classification of an image and object detection due to their deep architecture and its ability to learn spatial hierarchies.
 - **Inception Models:** Provide state-of-the-art performance in classification of an image tasks with improved efficiency.
 - **RNNs and LSTMs:** Effective for handling sequential and temporal data, making them ideal for tasks such as speech recognition and time-series prediction.
2. **Established Technologies:**
 - **CNNs and Inception Models:** Have been extensively researched and validated for a wide range of image processing tasks.

- **LSTM Networks:** Well-established for handling long-term dependencies in sequences.

Advantages of the Proposed System

1. **Gesture-Specific Analysis:**
 - **Hand Key Points:** Focuses on specific features relevant to gesture recognition, which can potentially improve the accuracy of gesture classification compared to general image recognition models.
2. **Real-Time Application:**
 - **Real-Time Gesture Recognition:** Enables immediate responses to detected gestures, crucial for applications like safety monitoring.
 - **Alerts and Notifications:** Provides a practical system for real-time incident notification, enhancing responsiveness to safety concerns.
3. **Adaptability and Innovation:**
 - **Hybrid and Stacking Models:** Leverages a combination of algorithms to potentially improve performance and robustness in gesture recognition tasks.
 - **Integration with MediaPipe and OpenCV:** Utilizes advanced tools for real-time video analysis and gesture detection, offering a modern approach to safety monitoring.

The existing system excels in established image and sequence analysis tasks with proven models like CNNs, RNNs, and LSTMs. The proposed system innovates by focusing on gesture recognition through hand key points, incorporating real-time processing, and integrating with alert systems. Each system has its strengths, with the existing system being well-suited for traditional image and sequence tasks and the proposed system offering novel applications for real-time safety monitoring.

VII. CONCLUSION

In this women's safety project, we successfully built a real-time gesture recognition system designed to enhance women's safety. By integrating MediaPipe for hand key point detection, various machine learning algorithms for gesture classification, and OpenCV for video processing, we created a system that accurately identifies sign language gestures and provides immediate alerts.

The project involved training and evaluating multiple machine learning models, including Decision Tree, Gaussian Naive Bayes, Gradient Boosting, and a hybrid Stacking Classifier. The selected model demonstrated robust

performance, meeting key evaluation metrics such as precision, recall, and F1-score.

Our approach not only enables precise gesture recognition but also facilitates real-time incident notification through automated alerts. This practical application underscores the system's effectiveness in responding swiftly to safety concerns.

Overall, the project showcases a novel and effective solution for real-time gesture recognition, with potential for further enhancements and applications in various fields.

VIII. REFERENCES

- [1] S. Simske, "Introduction, overview, and applications," *Meta-analytics*, 2019, doi: 10.1016/b978-0-12-814623-1.00001-0.
- [2] A.A. Barbhuiya, R.K. Karsh, and R. Jain, "CNN based feature extraction and classification for sign language," *Multimedia Tools and Applications*, vol. 80, no. 2, pp. 3051-3069, 2021.
- [3] P. Xanthopoulos and T. Razzaghi, "A weighted support vector machine method for control chart pattern recognition," *Computers and Industrial Engineering*, vol. 70, pp. 134-149, 2020.
- [4] S. Tornay, M. Razavi, and M. Magimai-Doss, "Towards multilingual sign language recognition," *Proceedings of the ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 6309-6313, 2020.
- [5] S.B. Carneiro, E.D.F.D.M. De Santos, et al., "Static gestures recognition for Brazilian sign language with Kinect sensor," *Proceedings of IEEE Sensors*, 2023, doi: 10.1109/ICSENS.2016.7808522.
- [6] R. Verma and A. Dev, "Vision based hand gesture recognition using finite state machines and fuzzy logic," *Proceedings of the International Conference on Ultra Modern Telecommunications and Workshops*, 2022, doi: 10.1109/ICUMT.2009.534542
- [7] A. Duarte, S. Palaskar, L. Ventura, D. Ghadiyaram, K. DeHaan, F. Metze, J. Torres, and X. Giro-i-Nieto. "How2Sign: A Large-scale Multimodal Dataset for Continuous American Sign Language". In: Conference on Computer Vision and Pattern Recognition (CVPR). 2021
- [8] Christo Ananth, P. Tamilselvi, S. Agnes Joshy, T. Ananth Kumar, "Blood Cancer Detection with Microscopic Images Using Machine Learning", *Machine Learning in Information and Communication Technology, Lecture Notes in Networks and Systems 498*, ISBN: 978-981-19-5090-2, pp. 45-54.
- [9] A. Graves, A.-r. Mohamed, and G. E. Hinton. "Speech Recognition with Deep Recurrent Neural Networks". In: CoRR abs/1303.5778 (2013). arXiv: 1303 . 5778. url: [http : //arxiv.org/abs/1303.5778](http://arxiv.org/abs/1303.5778).
- [10] Christo Ananth, M. Kameswari, R. Srinivasan, S. Surya, and T. Ananth Kumar, "A Novel Low-cost Visual Aid System for the Completely Blind People", *Machine Learning in Information and Communication Technology, Lecture Notes in Networks and Systems 498*, ISBN: 978-981-19-5090-2, pp. 177-183.
- [11] D. Uebersax, J. Gall, M. van den Bergh, and L. V. Gool. "Real-time sign language letter and word recognition from depth data". In: 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (2011), pp. 383–390.