Advanced International Journal for Research (AIJFR)
E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Smart Talk: Talk with Disable Person Without Any Barrier

Sankshipta Vaidya[1], Vedant Ambagade[2], Amol Rathod[3], Bhumika Tupe[4]

[1,2,3,4] Electronics & Telecommunication Engineering, S. B. Jain Institute of Technology, Management& Research, Nagpur.

Abstract

Communication difficulties present great obstacles to daily life for those who are unable to hear or speak. In this paper, we present an integrated assistive communication approach named as "Smart Talking: Talk with Disabled Person" to aid the interaction process eagerly of disabled community to the society. It comprises three sophisticated modules: the Sign Language Recognition, Emotion Detection, and Person Detection with Alert System. Based on computer vision, deep learning and natural language processing, the technique interprets hand gestures, detects emotional states and guarantees users' safety thanks to presence awareness. The experimental results show superior performance in gesture action recognition and emotion prediction. Such research is part of the development of intelligent assistive tools that can communicate in real-time, perceive emotions and constantly monitor safety.

Keywords: SLR, Emotion Detection, Person Detection, Deep Learning, CNN, Computer Vision, Assistive Technology Twilio AlertsHuman – Computer Interaction.

1. Introduction

Communication is the most fundamental aspect of human interaction, enabling the exchange of ideas, emotions, and information.

Yet, for millions of people around the globe who have speech, hearing or mobility issues — it can be a daily struggle to communicate. It's estimated by the World Health Organization (WHO) that more than 1.5 billion people live with some form of hearing loss, and almost 70 million use sign language as their first language. However, despite the high incidence of these conditions, a divide between disabled and other individuals remain, with no such general and inclusive communication method developed.

Conventional means of communication for differently abled people rely upon sign language which is an effective visual language using hand signed, facial expressions and gestures. Nevertheless, one significant drawback is that only a small minority of the population knows sign language, which limits communication between deaf persons and the rest of the world. This lack of communication can result in isolation, reliance, minimal opportunity for education, health, and employment.

The last few years the development of Artificial Intelligence (AI), Computer Vision and Deep Learning technologies has opened a wide-array of opportunities for assistive systems for communicating.

Solutions powered by AI can now interpret gestures, knowledge facial emotions and even detect human presence — in real time. These advances allow the development of intelligent systems able to "speak" for disabled people, understand emotion or guarantee safety in a wide range of contexts.

The proposed "Smart Talking: Talk with Disabled Person" system is a novel assistive communication system that combines three smart subsystems in one environment:

Sign Language Recognition Module:

This module translates gestures (hand movements) which are tracked by a camera, to text and speech. It allows those who suffer from speech impairments to communicate with ease without needing others to understand sign language. The architecture follows a combination of Convolutional Neural Networks (CNNs) and MediaPipe Hand Tracking for precise gesture classification.

Emotion Detection Module:

Speech is not only words but feelings. This service is a library that can analyze images to detect expressions of the face in real time and understand feelings such as entertaining, shocking, happy etc. Emotion recognition enables carers or listeners to perceive the emotional content of what is said.

Person Detection and Alert System:

Those who are disabled or elderly may need to be in a prison safe room with supervision. This module tracks the presence of a person in the space using face detection and reports when they are away too long by sending an alert via Twilio SMS or Telegram. It's a protective measure for people who live on their own.

All these modules together constitute a complete real-time communication/examination system for enabling the specially abled users. Through integration of gesture recognition, emotional intelligence, and automated monitoring functions the package is designed to give both parents and their children a holistic assistive experience which can reduce communication discrepancies, ensuring safety.

The project will alleviate three main challenges experienced by persons with disabilities:

Language barriers: Interpreting gestures into speech for greater comprehension.

Emotional Disconnect: Understanding and emotionalising the human's emotions appropriately.

Safety and risk: Yes, with real-time monitoring and alarm devices.

2. Literature Review

1.1 Historical Context and Evolution of Assistive Technologies

Assistive technologies for speech and hearing impairments have undergone considerable transformation in character over the past decades. It is well remembered that, conventionally, communication for this group heavily depended on manual sign language, written communications, and other mechanical devices of a rudimentary nature; there were, therefore, considerable barriers to social integration and independent functioning. The technological development in this area can be demarcated into three broad phases:

Pre-Digital Era (Before the 1980s)

Communication support tools were very minimal during this time. People mostly relied on manual sign languages, lip reading, and writing notes. Devices available then, like simple hearing aids, did not offer good amplification and were usually inaccessible. Social stigma and a non-supportive environment further restricted effective communication, making a person rely on a caregiver or interpreter.

Early Digital Revolution (1980s–2000s)

Digital electronics catalyzed the development of assistive communication. Early text-based communication devices could be used to type messages for audio output. Basic speech synthesizers were developed, but with restricted vocabularies and robotic articulation. Initial attempts with gesture recognition utilized simple sensors; these systems were not robust and were unsuitable for real-time natural communication settings. Modern AI Era (2010s–Present) Advances in computer vision, machine learning, and embedded computing have created state-of-the-art assistive technologies. Applications of neural networks in real-time sign language recognition become possible. Multi-modal communication that combines vision, audio, emotion detection, and contextual awareness has expanded support for people with impairments of their ability to communicate. Solutions rely on the cloud to update models and enable accessibility remotely, democratizing access to advanced assistive tools. The proposed Smart Talk System follows this modern trajectory, which merges real-time sign language recognition, emotion detection, and human presence monitoring into one integral solution. By incorporating all these functionalities into a lightweight AI pipeline, the system addresses long-standing challenges in accessible, intuitive, and contextually aware communication technologies.

Existing research in assistive communication technologies has shown significant progress in **sign language recognition**, **emotion detection**, and **ambient intelligence**. However, most studies have focused on these areas in isolation, resulting in gaps related to real-time integration, contextual understanding, and accessibility. Vision-based sign language recognition systems, such as the work by Li et al. (2020), have demonstrated high accuracy using advanced deep learning models like CNN-LSTM architectures. Their system achieved 94.2% accuracy for real-time American Sign Language (ASL) recognition, proving the feasibility of automated sign interpretation. Despite these advances, the approach was computationally intensive, limited to controlled lighting environments, and did not incorporate emotional context, which is crucial for natural communication.

In contrast, emotion recognition research by Wang and Chen (2019) focused on detecting human emotions through facial expression analysis using CNN-based ensemble methods. Their model achieved state-of-the-art accuracy on benchmark datasets such as FER2013 and AffectNet. However, it was restricted to static image analysis and lacked real-time adaptability or integration with other modalities like sign language—limiting its applicability in interactive, dynamic communication settings.

The field of ambient intelligence and assistive technology, explored by Smith et al. (2021), introduced frameworks for context-aware systems using multiple environmental sensors. Their work contributed to the development of privacy-preserving assistive environments, enhancing user safety and independence.

Nevertheless, the system required complex hardware setups and focused primarily on environmental monitoring rather than direct communication support.

3. METHODOLOGY

a. Proposed Solution

The Smart Talking Assistive System is designed to bridge communication gaps for the speech and hearing impaired by integrating multiple intelligent parameters into a single, cohesive platform. The system combines Real-Time Computer Vision, Deep Learning models, and SMS-based alert mechanisms to provide a holistic communication and security solution.

Users interact with the system through a standard webcam. The system simultaneously analyzes the video feed to interpret hand signs, understand the user's emotional state, and verify the presence of an authorized person. The translated text and emotional context are displayed in real-time, empowering users to communicate more effectively. If an unauthorized or unknown person is detected, an automatic alert is sent to a registered phone number, adding a layer of security.

b. Problem Identification

Many individuals, particularly those in the deaf and hard-of-hearing community, face significant barriers in daily communication. These challenges include: Isolation of Communication Tools: Existing solutions often focus only on translating sign language to text, completely ignoring the critical layer of emotional expression, which is a fundamental part of human interaction. Lack of Contextual Awareness: A text-based translation of "I am fine" can be misinterpreted without knowing if the user looks happy, sad, or neutral. This emotional disconnect can lead to misunderstandings. Vulnerability and Security Concerns: In certain environments (e.g., a smart home or a private workstation), there is no automated way to be alerted if an unrecognized person is present, which can be a security risk. Technological Barriers: Complex and expensive hardware setups (like specialized gloves) make assistive technology inaccessible to many. To address these problems, the Smart Talking system provides an integrated, visionbased platform that not only translates signs but also adds emotional context and ensures environmental awareness through person detection.

2.3 Functional Requirement

Accept real-time video input from a standard webcam.

Parameter 1: Sign Language Recognition: Detect and track hand movements/gestures and accurately convert them into textual output in real-time.

Parameter 2: Emotion Detection: Detect and classify the user's facial emotion into one of the four categories: Happy, Sad, Surprised, or Neutral.

Parameter 3: Person Detection: Continuously monitor the video feed to detect the presence of a human form. Integrate the outputs from all three parameters and display them on a single, userfriendly interface. Implement an alert system that automatically sends an SMS to a pre-registered phone number if no person is detected for a predefined continuous period. The system must operate with low latency to facilitate near real-time communication.

2.4 Technical Requirement

Component	Tools/TechnologiesUsed
Frontend	HTML,CSS,JavaScript
Backend	Python(core logic),flask
Machine Learning	TensorFlow/Keras(CNN Model),MediaPipe/Cvzone,OpenCv.
Sign Language Detection	CNN Classifier (Sign Alphabets), MediaPipe Hand Landmarks
Emotion Recognition	CNN Model, Haarcascade Front-face Detector
Person Detection	OpenCV (Haar Cascade), Twilio API (SMS Alert), Telegram Bot API
Data Storage	Local JSON/CSV files for logging detected sentences, alerts, predictions
Messaging / Alerts	Twilio SMS API, Telegram Bot Commands

2.5 System Work Flow

The Smart Talk system operates through a unified workflow that integrates sign language recognition, emotion detection, and person-presence monitoring in real time. The process begins with continuous video input captured from a webcam, which is simultaneously analyzed for hand gestures, facial expressions, and user presence. For sign language recognition, the hand region is detected using MediaPipe landmarks, preprocessed, and then classified through a trained CNN model to convert gestures into text and speech. Emotion detection extracts the face area using Haarcascade, processes it into a 48×48 grayscale frame, and uses a CNN classifier to identify the user's emotional state. In parallel, the presence detection module continuously checks for a face in the frame; if no person is detected for a predefined duration, the system triggers an automated alert through Twilio SMS and allows remote control via Telegram Bot. The outputs from all three modules—recognized signs, detected emotions, and safety alerts—are displayed on a single user interface, enabling smooth, real-time communication without barriers.

3. SYSTEM ARCHITECTURE

3.1 System Design and Management

The Smart Talk Assistive System is developed with a user-centric and modular design approach, ensuring that communication and security are both seamless and accessible for speech and hearing-impaired individuals. The architecture focuses on real-time performance, accuracy, and modularity, making it suitable for both current use and future upgrades, such as expanding the sign language vocabulary or adding new emotional states.

3.1.1 System Architecture Overview

The system is designed using a multi-threaded, pipeline architecture where each core module operates in parallel to ensure real-time performance and a responsive user experience. The flow of data is managed centrally to synchronize the outputs from the three independent parameters.

1. Frontend Layer (User Interface)

The Frontend Layer serves as the primary visual interface between the user and the system. It is developed using HTML, CSS, and JavaScript to create a clean, intuitive, and real-time display. This layer is responsible for rendering the video feed and overlaying the system's analysis.

Key Features:

- Displays a real-time video stream from the user's webcam.
- Shows a dynamic bounding box around the user's hands and face.
- Presents the translated sign language text in a large, clear font.
- Displays a dynamic emoji or label for the detected emotion (e.g., Happy).
- Indicates the status of person detection (e.g., "Person Detected" or "No Person").

2. Backend Layer (Core Processing Engine)

The Backend Layer, built using the Flask framework (Python), acts as the central processing unit of the system. It handles all server-side logic, model inference, and data flowcoordination. It utilizes multi-threading to run the three parameter modules concurrently without blocking the main interface.

Major Responsibilities:

- Captures and preprocesses the video feed from the webcam using OpenCV.
- Manages three parallel processing threads:
  - Model 1 (Sign Interpreter): Executes the hand landmark detection (MediaPipe) and
  - Model 2 (Emotion Analyzer): Executes the face detection (Haar Cascade/MediaPipe) and emotion classification model (CNN).
  - Model 3 (Security Monitor): Executes the person detection model (YOLO) and manages the alert timer logic.
- Aggregates the results from all threads and sends them to the frontend via a web socket or a continuous HTTP stream.
- Hosts the SMS Alert API endpoint, which is triggered by the Security Monitor to send alerts via the

Twilio API.

3. AI & Computer Vision Layer

This layer contains the pre-trained machine learning and deep learning models that form the intelligence of the system. It is tightly integrated with the backend processing logic. Core Components: • Sign Language Model: A Convolutional Neural Network (CNN) trained on a dataset of hand gestures, capable of classifying static or dynamic signs into textual output. • Emotion Recognition Model: A CNN trained on the FER2013 or similar dataset, specialized in classifying cropped facial images into the four emotions: Happy, Sad, Surprised, and Neutral. • Person Detection Model: A pre-trained YOLO (You Only Look Once) model, optimized for fast and accurate real-time object detection, specifically tuned to identify "person" classes.

4. Data & Service Integration Layer

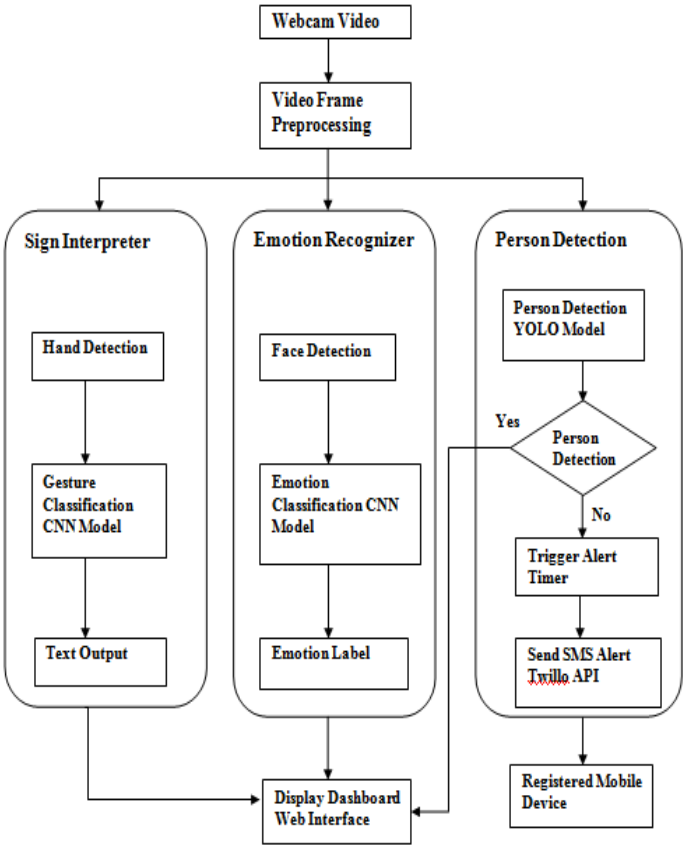

Figure 4.1: System Block diagram of Smart Talking

This layer handles all external communications and data persistence for the system. Key Integrations: • SMS Gateway (Twilio API): Provides the service for sending instant SMS alerts to the pre-registered phone number when the security condition is met.

This diagram illustrates the three-parameter architecture of your system.

1. Input Layer: A single Webcam Video Feed serves as the input for all three processing modules.
2. Core Processing Engine: The video feed is preprocessed and then simultaneously processed by three parallel pipelines:
- **Sign Interpreter:** Detects hands, classifies the gesture using a CNN model, and generates a Text Output.
- **Emotion Recognizer:** Detects the face, classifies the emotion using another CNN model, and generates an Emotion Label (Happy, Sad, Surprise, Neutral).
- **Person Detection:** Detects if a person is present using a YOLO model. If no person is detected for a continuous period, it triggers an SMS Alert to a registered mobile number via an API.
3. Output & Interface Layer: All results are aggregated and displayed on a central Display Dashboard (a web interface). The SMS Alert is sent directly to a mobile device, providing a separate security notification channel.

4. DATA ANALYSIS

The Smart Talk Assistive System uses data collected from real-time video input to analyze and interpret user communication and context through three parallel parameters. This process involves several data analysis and machine learning steps to ensure accuracy, real-time performance, and robustness.

1. Data Collection:

- For Sign Language Recognition: A dataset of hand gesture images or sequences wascollected or utilized. This dataset includes various signs (e.g., A-Z), with each data point containing the image frames and the corresponding textuallabel.
- For Emotion Detection: A standard dataset such as FER2013 was used, which containsthousands of gray scale, cropped facial images labeled with seven core emotions. For thisproject, the model was tailored to classify four of them: Happy, Sad, Surprise, and Neutral.
- For Person Detection: A large-scale image dataset with bounding box annotations forthe "person" class (e.g., COCO dataset) was used to pre-train the YOLO model, enabling
it to robustly detect human forms in various poses and environments.

2. Data Preprocessing:

- Image Standardization: All input video frames are resized to a fixed dimension (e.g.,224x224 pixels) and normalized to have pixel values between 0 and 1 to facilitate stable model training and inference.
- For Sign Language: Hand regions are cropped and isolated. Techniques likebackground subtraction or landmark detection (using MediaPipe) are employed to focuson the relevant features, reducing computational noise.
- For Emotion Detection: Faces are detected and cropped from the full frame. Imagesare converted to grayscale as color is not a critical feature for emotion classification, simplifying the model.
- Data Augmentation: To improve model generalization, techniques like rotation,scaling, and brightness adjustments were applied to the training data for both signlanguage and emotion models, making them invariant to minor real-world variations.

3. Feature Extraction & Model Training:
- Sign Language & Emotion Models: Convolutional Neural Networks (CNNs) are usedfor automatic feature extraction and classification. The CNNs learn hierarchicalpatterns—from simple edges in initial layers to complex shapes and textures in deeperlayers—directly from the pixel data, eliminating the need for manual feature engineering.
- Person Detection Model: The YOLO (You Only Look Once) architecture is used. Itis a single, unified model that simultaneously predicts bounding boxes and classprobabilities for those boxes, making it extremely fast and accurate for real-time objectdetection.

4. Real-Time Inference and Integration:
- Parallel Processing: In the live system, preprocessed video frames are fed27simultaneously into the three trained models.
- Output Aggregation: The system aggregates the output from all three pipelines:
- The classified text from the Sign Language Model.
- The emotion label from the Emotion Recognition Model.
- The binary status (Person Detected/Not Detected) from the Person DetectionModel.
- Decision Logic: The Person Detection module runs a continuous check. If the "NotDetected" state persists beyond a predefined timer threshold, it triggers the alertmechanism.

5. Alert Mechanism:
- The system integrates with the Twilio SMS API. When the security logic is triggered,a pre-defined alert message is automatically sent to a registered phone number providingan immediate external notification. The performance of this module is measured by itsreliability and speed in sending the alert.

5. Conclusion

The SMART TALK: Talk with Disabled Person Without Any Barrier project is a comprehensive solution designed to address several critical challenges faced by disabled individuals in their daily lives. Disabled persons often experience difficulties in communicating with others, expressing their emotions, and ensuring personal safety. Traditional systems and tools generally focus on one of these aspects at a time, such as communication aids or monitoring devices, but fail to provide an integrated solution. This limitation often forces disabled individuals to depend on multiple devices or the constant support of caregivers, which can reduce their independence, confidence, and dignity.

The SMART TALK system has been developed to overcome these challenges bycombining three essential modules: Sign Language Recognition, EmotionDetection, and Person Detection. Each module contributes uniquely to making lifeeasier, safer, and more inclusive for disabled individuals.The Sign Language Recognition module is a major breakthrough for communication.Many disabled persons rely on sign language to interact but they often face difficultiescommunicating with people who do not understand it. By translating hand gestures intoreadable text and audible speech in real-time, this system removes the language barrier,allowing seamless communication between disabled individuals and others. This notonly enhances independence but also encourages social interaction and inclusion indaily activities, whether at home, in educational settings, or in public spaces.The Emotion Detection module adds an important

dimension of emotionalunderstanding. Often, the emotional state of a disabled person goes unnoticed becausenon-verbal cues may not be easily interpretable. By analyzing facial expressions andidentifying emotions such as happiness, sadness, anger, or surprise, the system enablescaregivers, family members, and friends to understand and respond appropriately. Thisimproves emotional connection, provides timely support, and fosters a more empatheticenvironment. Users feel understood, cared for, and socially connected, which is crucialfor mental well-being and overall quality of life.

The Person Detection feature ensures safety and situational awareness. The system monitors the presence of the individual and responds accordingly. This is particularly helpful for caregivers to know if the person is in the monitored area and provides assurance of supervision without being intrusive. It also allows the system to function intelligently, triggering alerts or guidance when necessary. By including this module, the project combines communication and emotional support with a basic level of safety, making it a truly multi-functional system.

A key strength of SMART TALKING is its user-friendly mobile application interface, which integrates all three modules into a single platform. The interface is designed to be simple, intuitive, and accessible, allowing disabled persons and their caregivers to use the system without extensive technical knowledge. All functions, such as translating sign language, detecting emotions, and monitoring presence, can be accessed easily, making the system practical for real-world daily use. Unlike existing solutions that operate in isolation, this project unifies multiple functionalities in one platform, reducing complexity and increasing effectiveness. The development of SMART TALKING highlights the potential of technology to empower disabled individuals. It supports autonomy by allowing users to communicate independently, express emotions freely, and stay safe with minimal external supervision. Furthermore, the system fosters inclusion by bridging the communication gap between disabled and non-disabled individuals, helping them participate more actively in social, educational, and professional environments.

In addition to its current capabilities, SMART TALKING has significant potential for future improvements. For example, remote caregiver notifications could be added to alert family members when attention is needed. The system could support multi-user environments, such as classrooms or group activities, to provide real-time monitoring and communication support for multiple disabled persons simultaneously. Advanced personalization using AI could adapt the system based on individual user behavior, preferences, and communication style, making interactions even more natural and effective. Integration with other assistive devices or smart home systems could further enhance safety, convenience, and independence for disabled individuals.

In conclusion, SMART TALKING: Talk with Disabled Person Without Any Barrier is more than just a technological solution. It represents a step toward a more inclusive, empathetic, and supportive society. The project empowers disabled individuals to communicate freely, express emotions, and maintain safety, enhancing their confidence, independence, and quality of life. It serves as a foundation for future innovations in assistive technology, inspiring more inclusive solutions that bridge gaps between people with disabilities and the world around them. With continued development and user-centric improvements, SMART TALKING has the potential to become a comprehensive, life-enhancing tool that truly breaks the barriers faced by disabled individuals.

6. FUTURESCOPE

The SMART TALKING system has been designed to provide communication support, emotional understanding, and person detection for disabled individuals. While the current implementation addresses these critical needs effectively, there is a significant potential to expand its functionality, improve usability, and adapt it to a wider range of scenarios. The further scope of this project includes:

1. Advanced Personalization
   - The system can be enhanced to learn individual user preferences, communication styles, and emotional patterns over time.
   - By incorporating adaptive algorithms, the app could provide personalized suggestions, communication assistance, and emotion-based guidance tailored to each user.

2. Integration with Smart Devices
   - The system can be connected with wearable devices, smart watches, or home automation systems to provide real-time monitoring and assistance.
   - For example, detecting unusual activity through smart sensors or sending alerts to caregivers if the user is in a risky situation.

3. Multi-Language Sign Language Support
   - Currently, the app may support a single sign language system. Expanding it to include multiple regional or international sign languages would help reach a broader user base.
   - This would allow communication across different communities, making the platform globally accessible.

4. Remote Monitoring and Alerts
   - Caregivers or family members could receive live notifications and status updates through the app, improving safety and responsiveness.
   - Features like video streaming, instant alerts, or emergency calling can be incorporated to handle urgent situations effectively.

5. Enhanced Emotion Detection
   - Emotion detection could be extended to recognize more nuanced emotional states, stress levels, or mental health indicators.
   - Integration with counseling or therapy suggestions can be included to provide timely psychological support.

6. Multi-User Environment Support.
   - The platform can be expanded to support classrooms, group activities, or workplaces, allowing multiple users to interact simultaneously with realtime assistance.
   - This would make the app useful in educational institutions and community centers.

7. Offline Functionality
   - Certain core features, such as sign language translation and person detection, can be made to work offline.

- This would ensure accessibility even in areas with poor internet connectivity.

References

1. Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2016). "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild." IEEE Transactions on Affective Computing, 10(1), 18-31.
2. Mekruksavanich, S. & Jitpattanakul, A. (2021). “A Smartphone-Based Online System for Fall Detection with Alert Notifications.”
3. Kumar, A., Prakash, V., & Gupta, S. (2018). "Indian Sign Language Recognition System Using Convolutional Neural Networks." Proceedings of the 2nd International Conference on Communication and Electronics Systems (ICCES), 465-469.
4. Waiswa, N., & Namirembe, S. (2020). "Assistive Technologies and Their Role in Healthcare for the Differently Abled." International Journal of Health Sciences and Research, 10(6), 11-17.
5. Haque, M. A., Irfan, M., Hossain, M. S., & Rahman, M. A. (2019). Emotion Recognition Using Deep Learning Techniques: A Review. Journal of Machine Learning Research (JMLR).
6. Koller, O., Forster, J., & Ney, H. (2015). Continuous Sign Language Recognition: Towards Large Vocabulary Statistical Recognition Systems Handling Multiple Signers. IEEE Transactions on Human-Machine Systems.
7. Aziz, O., Musngi, M., Park, E. J., Mori, G., & Robinovitch, S. N. (2017). A Comparison of Accuracy of Fall Detection Algorithms (Threshold-Based vs. Machine Learning) Using Wearable Sensors. Sensors (MDPI)