A Survey on Kannada Sign Language Recognition Using Deep Learning

Harshitha G J[1], Dr. Manjunatha Kumar B H[2]
[1,2]*Department of Computer Science and Engineering, S J C Institute of Technology, Chikkaballapura*

Abstract*—Sign language recognition has emerged as a crucial research area in assistive computing, aiming to bridge the communication gap between hearing-impaired individuals and the general population. For individuals with hearing and speech disabilities, sign language serves as the primary means of communication, supporting effective interaction in daily life. However, the absence of widespread sign language knowledge among the general population often leads to communication difficulties, emphasizing the need for automated and intelligent translation systems. Although significant research has been conducted on American Sign Language (ASL) and Indian Sign Language (ISL), regional sign languages such as Kannada Sign Language (KSL) remain largely unexplored. Recent advancements in computer vision and deep learning have enabled the development of camera-based sign language recognition systems that are non-intrusive, cost-effective, and suitable for real-time deployment. This under representation is mainly due to the lack of standardized datasets, limited linguistic resources, and insufficient region-specific recognition frameworks. Consequently, existing models often fail to generalize well to KSL, highlighting the necessity for focused research in this area. Image-based Convolutional Neural Networks (CNNs) are widely used for extracting spatial features from hand gestures, while landmark-based approaches using MediaPipe provide accurate hand keypoint detection with reduced computational complexity. Additionally, sequential models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks effectively capture temporal variations in dynamic gestures, and transformer-based methods offer improved modeling of complex gesture patterns. This literature survey reviews and compares these approaches, identifies key challenges, and highlights research gaps, thereby motivating the development of a real-time, deep learning-driven Kannada Sign Language recognition system. Such a system can enhance accessibility, promote inclusive communication, and support practical applications in education, public services, and human– computer interaction. By focusing on accuracy, efficiency, and real-time performance, future research can significantly improve the quality of life for the hearing-impaired community.

Index Terms*—Kannada Sign Language, Deep Learning, MediaPipe, CNN, LSTM, Gesture Recognition, Assistive Systems.

I. INTRODUCTION

Communication plays a vital role in human interaction, influencing social relationships, education, professional growth, and access to essential public services. For individuals with hearing and speech impairments, sign language acts as the primary medium of communication, enabling them to express ideas, emotions, and information effectively. In India, sign language is not uniform; it varies across regions due to differences in language, culture, and social practices. Kannada Sign Language (KSL) is commonly used by the hearing- impaired community in Karnataka. However, limited familiarity with KSL among the general population continues to create significant communication challenges.

At present, interaction between sign language users and non-signers largely depends on human interpreters. While interpreters are essential for accurate communication, their availability is often limited, particularly in everyday or emergency situations. Moreover, reliance on interpreters restricts spontaneous communication and continuous interaction. These challenges have encouraged the development of automated sign language recognition systems that can translate hand gestures into understandable text or speech in real time, promoting independence for the hearing-impaired community. Initial research in sign language recognition primarily focused on hardware- based solutions such as data gloves, motion sensors, and depth cameras.

Although these systems provided precise gesture measurements, they were costly, uncomfortable to use, and impractical for daily applications. With advancements in computer vision and deep learning, camera-based recognition systems have gained popularity. These vision-based approaches use standard cameras to capture hand gestures, making them affordable, non-intrusive, and suitable for real-world environments. Despite significant progress in this field, most research efforts have concentrated on American Sign Language (ASL) or general Indian Sign Language (ISL).

Regional sign languages such as Kannada Sign Language remain insufficiently explored due to the lack of publicly available datasets and region-specific recognition frameworks. This literature survey aims to review existing sign language recognition techniques, research gaps, thereby providing a foundation in concern of developing an effective deep learning–based Kannada Sign Language recognition system. By addressing these gaps, future research can contribute to inclusive assistive technologies that support regional languages. An efficient KSL recognition system can enhance accessibility in education, healthcare, and public services, enabling seamless interaction between signers and non-signers in everyday communication scenarios.

1.1. BACKGROUND OF SIGN LANGUAGE COMMUNICATION

Sign language is a complete and natural form of human communication that uses hand gestures, facial expressions, and body movements to convey meaning. For individuals with hearing and speech impairments, it serves as the primary medium for expressing ideas, emotions, and intentions, enabling participation in social, educational, and professional activities. Unlike spoken languages, sign languages are visual–gestural in nature and possess their own grammar, syntax, and vocabulary, making them independent linguistic systems rather than simple representations of spoken language.

In India, sign language communication is highly diverse due to regional, cultural, and linguistic variations. Although Indian Sign Language (ISL) provides a broad framework, many regions have developed localized signing practices that better reflect their spoken languages. Kannada Sign Language (KSL) is one such regional variant widely used by the hearing-impaired community in Karnataka. KSL incorporates gestures and expressions aligned with the Kannada language, making it more intuitive for native users. However, the lack of standardization and limited exposure among non-signers significantly restrict effective communication between the hearing-impaired community and the general population.

Traditionally, sign language communication in public spaces, educational institutions, healthcare centers, and workplaces depends on trained human interpreters. While interpreters play a crucial role in facilitating understanding, their availability is limited, and they cannot support continuous or spontaneous interaction in all situations. This reliance often leads to communication delays, reduced independence, and limited accessibility for sign language users.

The evolution of digital technologies has opened new possibilities for enhancing sign language communication. Advances in computer vision, machine learning, and deep learning have enabled the development of automated sign language recognition systems that can interpret gestures using camera-based inputs. These systems aim to translate signs into readable text or speech in real time, reducing dependency on interpreters. However, most existing solutions focus on widely studied sign languages, leaving regional languages like Kannada underrepresented. This background highlights the need for research-driven, technology-enabled solutions that support regional sign language communication and promote inclusivity through accessible and intelligent systems.

1.2. KANNADA SIGN LANGUAGE AND ITS IMPORTANCE

Kannada Sign Language (KSL) is a regional sign language used by the hearing- and speech-impaired community in the state of Karnataka. It has evolved naturally over time to reflect the linguistic structure, cultural context, and daily communication needs of Kannada-speaking individuals. Like other sign languages, KSL uses a combination of hand shapes, hand movements, orientation, facial expressions, and body posture to convey meaning. It is not a direct translation of spoken Kannada but an independent visual language that enables effective communication among its users.

Although Kannada Sign Language is widely used by the deaf community in Karnataka, it is not well understood by the general population. In places such as schools, hospitals, government offices, and workplaces, there is very little support for KSL users, which creates serious communication difficulties. As a result, hearing- impaired individuals often depend on family members or human interpreters, reducing their independence and limiting access to essential services. Unlike existing systems that mainly support American or generic Indian Sign Language, the proposed system captures the unique gestures and linguistic features of KSL, ensuring better accuracy, cultural relevance, and inclusive communication for Kannada-speaking users. In the proposed project, Kannada Sign Language serves as the core focus for data collection, model training, and system evaluation. By leveraging computer vision, MediaPipe-based hand landmark detection, and deep learning models such as CNNs, the system is designed to accurately recognize KSL hand gestures and convert them into readable Kannada text in real time. This approach not only enhances accessibility but also promotes digital inclusion for the hearing-impaired community in Karnataka. Furthermore, the project lays a strong foundation for future extensions, including dynamic gesture recognition, sentence formation, and speech synthesis, thereby contributing to the broader goal of inclusive and assistive communication technologies.

1.3. MOTIVATION FOR AUTOMATED SIGN LANGUAGE RECOGNITION

The motivation for automated sign language recognition stems from the growing need to ensure equal communication access for individuals with hearing and speech impairments. In many real-world scenarios, such as classrooms, hospitals, government offices, and workplaces, effective communication is essential for safety, learning, and participation.

However, the limited availability of trained sign language interpreters often forces hearing-impaired individuals to depend on others, reducing their independence and privacy. This challenge is more pronounced for users of regional sign languages like kannada sign Language (KSL), which are rarely supported by existing technological solutions

Although recent advancements in artificial intelligence have improved sign language recognition systems, most research and commercial tools focus on American Sign Language or generalized Indian Sign Language. These systems fail to capture the linguistic variations, gesture patterns, and cultural context unique to KSL, making them less effective for Kannada-speaking users. As a result, there is a clear need for a region-specific automated solution that can accurately recognize KSL gestures.

The rapid development of computer vision, deep learning, and hand landmark detection technologies provides a strong foundation for building such systems. Camera-based approaches using MediaPipe and convolutional neural networks allow real-time gesture recognition without the need for expensive or intrusive hardware. Automating Kannada Sign Language recognition not only improves accessibility but also promotes digital inclusion by enabling seamless interaction between signers and non-signers. This project is motivated by the goal of developing a scalable, user-friendly, and cost-effective assistive communication system that empowers the hearing-impaired community and supports inclusive societal participation.

II. RELATED WORK ON SIGN LANGUAGE RECOGNITION

With the rapid growth of computer vision and artificial intelligence, sign language recognition has emerged as an important research area aimed at improving communication accessibility for hearing- and speech-impaired individuals. Vision-based sign language recognition systems have gained popularity due to their non-intrusive nature and ease of deployment using standard cameras. However, recognizing regional sign languages such as Kannada Sign Language (KSL) remains challenging due to limited datasets, signer variations, and environmental constraints.

Early studies on KSL primarily targeted static gestures. Saraswathi et al. [1] proposed a hybrid machine learning framework combining MediaPipe-based hand landmark extraction with MobileNetV2 for static gestures and 3D- CNN for dynamic gestures. Although the system achieved real-time performance on mobile devices, its accuracy dropped under varying lighting conditions or partial occlusions, highlighting the need for robust feature extraction methods. Shruthi et al. [2] extended this work by using CNNs trained on MediaPipe- extracted key points instead of full-frame images. Their focus on hand landmarks improved classification accuracy for static gestures, but the approach was restricted to isolated gestures, making it unsuitable for sentence-level translation or continuous signing.

To broaden applicability, research has also targeted Indian Sign Language (ISL). which are rarely supported by existing technological solutions keypoint extraction, achieving robust real-time gesture recognition in complex environments. Despite improved detection, the increased computational requirements posed limitations for deployment on low-end devices. Hon et al. [4] explored CNN-only systems for static ISL gestures using camera inputs, highlighting simple and accessible solutions. However, their method struggled with dynamic sequences and background clutter, emphasizing the need for temporal modeling in gesture recognition systems.

Recognizing dynamic gestures, which involve continuous movement, is more complex due to temporal dependencies. Biswas et al. [5] combined MediaPipe landmark extraction with LSTM networks, effectively modeling temporal patterns to recognize dynamic gestures. Their approach demonstrated improved performance over CNN-only systems but required large, carefully annotated datasets for training. Subramanian et al. [7] further improved temporal modeling by employing optimized GRU architectures with MediaPipe landmarks, achieving faster convergence and higher stability. These studies indicate the importance of recurrent architectures in capturing sequential dependencies in sign language.

While deep learning dominates recent research, classical computer vision techniques still provide useful insights. Katoch et al. [8] combined SURF feature extraction with Bag of Visual Words and SVM/CNN classifiers for static ISL alphabets and numbers. Such hybrid approaches require less training data and can work effectively in resource-constrained scenarios, but they generally fail to handle dynamic gestures or continuous signing effectively.

Video-based approaches extend static recognition to continuous sequences. Puranik et al. [9] leveraged Recurrent Neural Networks (RNNs) to capture temporal correlations across video frames. This approach is effective for recognizing complex sentence-level gestures but demands large, annotated video datasets and is sensitive to camera motion, frame inconsistencies, and occlusion. Similarly, attention-based models have been explored to improve focus on critical gesture frames, mitigating some challenges of background clutter and signer variation.

Recently, transformer-based models have been applied to sign language recognition due to their superior feature representation capabilities. Nedungadi et al. [10] proposed a Vision Transformer-based model for ISL conversational agents, demonstrating robustness to variations in hand pose and background. However, transformer architectures are computationally intensive and require large datasets, limiting real-time deployment on embedded or mobile devices. Additionally, their applicability to regional languages such as KSL remains under-explored, partly due to dataset scarcity.

From the literature, it is evident that most existing systems either focus on static gesture recognition or require high computational resources for the dynamic gesture handling. Additionally, limited availability of a

standardized Kannada Sign Language datasets remains a major challenge. These gaps motivate the proposed system, which leverages MediaPipe-based hand landmark extraction combined with CNN-based classification to achieve accurate, real-time Kannada Sign Language recognition with reduced computational complexity and improved scalability.

III. RESEARCH GAPS AND CHALLENGES

Despite the growing interest in sign language recognition, significant research gaps persist, particularly for regional languages such as Kannada Sign Language (KSL). Most existing studies focus on widely used languages, including American Sign Language (ASL) or general Indian Sign Language (ISL), leaving regional sign languages underexplored. This has resulted in a scarcity of large, annotated datasets for KSL, which is a critical limitation. A robust dataset must include diverse participants, variations in hand size and shape, different lighting conditions, multiple backgrounds, and various camera angles. Without such diversity, recognition models struggle to generalize beyond the specific users or controlled environments they were trained on, limiting their effectiveness in real-world applications.

A major limitation across many current systems is the focus on static gestures. While these approaches work well for recognizing alphabets or isolated words, real-life communication often involves continuous gestures, combining multiple signs into phrases and sentences. Dynamic gesture recognition requires models to understand temporal dependencies, meaning the system must learn how gestures evolve over time and relate to each other contextually. Existing approaches that attempt to handle dynamic gestures often require carefully segmented sequences and large annotated datasets, which are largely unavailable for KSL. This creates a significant barrier to building systems that support natural and fluent communication.

Environmental variability introduces additional challenges. Real-world conditions such as changes in lighting, cluttered backgrounds, occlusions, motion blur, and camera angles can drastically reduce system accuracy. Landmark-based approaches, such as those using MediaPipe, partially address this problem by focusing on hand keypoints rather than full-frame images, thereby reducing background influence. However, these methods still struggle under scenarios involving fast hand movements, multiple signers, overlapping hands, or partial visibility, making consistent recognition difficult. Furthermore, inter-user variability—differences in hand size, shape, and signingstyle—introduces additional complexity, causing intra- class variations that degrade recognition performance if the model has not been trained on a sufficiently diverse dataset.

Computational efficiency and real-time performance remain critical concerns for practical deployment. While advanced deep learning models, such as 3D Convolutional Neural Networks (3D-CNNs) or Vision Transformers, achieve high recognition accuracy, they often require significant computational resources, limiting deployment on mobile devices, embedded systems, or resource-constrained environments. Lightweight models or landmark-based approaches reduce resource consumption but often come at the cost of reduced accuracy or slower convergence in complex or dynamic gesture scenarios. Balancing model accuracy, latency, and computational efficiency remains a pressing challenge for real-time applications.

Another significant limitation is scalability and adaptability. Most existing systems are trained on fixed gesture sets, making it difficult to add new gestures, regional variations, or sentence-level sequences without retraining the entire model. This lack of flexibility hampers the system's adaptability to evolving user needs or expanded communication scenarios. Furthermore, many systems provide minimal interactive feedback or guidance to the user, which reduces usability, particularly for applications in education, healthcare, or public services where real-time response and user engagement are essential.

Additionally, most existing KSL recognition systems focus primarily on hand gestures, with limited multimodal integration. Facial expressions, body posture, and head movements are critical for capturing the full semantic meaning of a gesture and are widely used in natural sign languages to convey emphasis, emotion, or grammar. Ignoring these components limits the system's semantic understanding, making it less effective for conversational or context-aware applications.

The study of Kannada Sign Language recognition highlights several critical research gaps. The absence of large, standardized datasets limits model generalization across users and environments. Most existing systems focus on static gestures, offering limited support for dynamic or sentence-level recognition required for natural communication. Performance is often affected by environmental variations such as lighting, background clutter, and occlusion, as well as user differences in hand shape and signing style. Although high-accuracy models exist, their computational demands restrict deployment on mobile or low-end devices. Additionally, many systems lack scalability, multimodal feature integration, and interactive user interfaces, reducing their practical usability. Addressing these challenges is essential for developing efficient, robust, and real-world applicable KSL recognition systems. By leveraging hand landmark detection, classification and pre-processing, future systems can achieve higher robustness, generalization.

Table 1. Summary of research gap, limitations and solution

Category	Description	Limitation	Solution
Dataset Scarcity	Limited availability of large, annotated KSL datasets representing diverse users and conditions.	Models fail to generalize to new users or real-world scenarios.	Create comprehensive datasets with multiple signers, backgrounds, and lighting conditions.
Dynamic Gesture Recognition	Most systems focus on static gestures; continuous gestures, phrases, and sentences are underexplored.	Real-life communication is not fully supported.	Use temporal models (LSTM, GRU, or hybrid CNN-RNN) to capture motion sequences.
Environmental Sensitivity	Recognition is affected by lighting, cluttered backgrounds, occlusion, and motion blur.	Reduced accuracy in uncontrolled, real-world environments.	Apply pre-processing, data augmentation, and landmark-based feature extraction (e.g., MediaPipe).
User Diversity	Variations in hand size, shape, and signing style across individuals.	Increased intra-class variability reduces model reliability.	Train models on diverse datasets with multiple signers and perform normalization techniques.
Computational Constraints	High-performing models (e.g., 3D-CNN, Vision Transformers) require substantial resources.	t suitable for mobile or embedded devices.	Develop lightweight CNN models or optimized hybrid architectures for real-time performance.
Scalability & Extensibility	Fixed gesture sets; adding new gestures or multi-gesture sequences often requires retraining.	Limited adaptability to new gestures or sentence-level translation.	Design modular, flexible pipelines with incremental learning or transfer learning capabilities.
Limited Multimodal Integration	Most systems focus only on hand gestures, ignoring facial expressions and body pose.	Incomplete semantic understanding of signs.	Incorporate holistic features (face, body, pose) alongside hand landmarks for richer representation.
Interface & Interaction Limitations	Lack of real-time feedback, guidance, or adaptive interfaces.	Reduced usability and difficulty for practical applications.	Develop interactive interfaces with live video guidance, real-time text display, and optional speech output.

IV. PROPOSED METHODOLOGY

4.1. SYSTEM OVERVIEW

The proposed system is a vision-based Kannada Sign Language (KSL) recognition framework aimed at enabling real-time communication for hearing- and speech-impaired individuals. The system captures live video input using a standard RGB camera and processes it through a modular pipeline consisting of data acquisition, pre-processing, hand landmark detection, deep learning-based classification, and real-time output generation. By using MediaPipe-based on hand landmark extraction and a lightweight deep learning model, the system achieves high accuracy while maintaining low computational complexity, making it suitable for real- world deployment on low-resource devices.

The architecture follows a modular pipeline consisting of data acquisition, pre-processing, hand landmark detection, feature extraction, gesture classification, and output generation. Each module operates independently while maintaining seamless integration with the others, ensuring scalability and ease of maintenance. This modularity allows new gestures, improved models, or additional modalities to be integrated without redesigning the entire system.

4.2. DATA ACQUISITION

Data acquisition is a critical component of the system, as the quality and diversity of training data directly influence recognition accuracy. Gesture data is collected using a standard RGB camera, capturing both images and video sequences of Kannada Sign Language gestures. The dataset includes samples from multiple users of different age groups and hand characteristics to capture inter-user variability.

Both static gestures, such as alphabets and isolated words, and dynamic gestures involving continuous hand movements are recorded. Data is collected under varying environmental conditions, including different lighting intensities, backgrounds, and camera angles, to improve robustness. Each gesture sample is manually labeled with its corresponding Kannada sign, ensuring accurate ground truth for supervised learning. To further enhance dataset diversity, multiple repetitions of each gesture are recorded from each participant. This approach helps the model learn consistent gesture patterns while also adapting to natural variations in gesture execution.

4.3. DATA PREPROCESSING

Pre-processing is essential to enhance data quality and ensure uniformity across all input samples. Raw image frames are resized to a fixed resolution to maintain consistency and reduce computational complexity. Pixel normalization is applied to reduce the influence of lighting variations and improve numerical stability during training.

Noise reduction techniques are used to suppress unwanted artifacts caused by camera noise or motion blur. For landmark-based features, pre-processing includes normalization of landmark coordinates relative to a reference point, typically the wrist joint. This ensures scale and translation invariance, allowing the system to recognize gestures regardless of the user's distance from the camera.

Data augmentation techniques such as rotation, scaling, translation, horizontal flipping, and brightness adjustment are applied to artificially expand the dataset. For dynamic gesture sequences, temporal smoothing and frame alignment techniques are used to reduce jitter and ensure consistent motion representation. These pre- processing steps significantly improve model robustness and generalization performance.

4.4. HAND LAND MARK DETECTION

Hand landmark detection is a crucial stage in the recognition pipeline, as it extracts meaningful structural information about the hand while eliminating irrelevant background details. The proposed system utilizes the MediaPipe Hands framework for real-time hand detection and tracking. MediaPipe identifies the hand region within each video frame and extracts 21 anatomically significant landmarks corresponding to the wrist, palm, and finger joints.

These landmarks provide a compact and expressive representation of hand posture and finger articulation. By tracking landmark positions across consecutive frames, the system captures both spatial configurations and temporal motion patterns, which are essential for recognizing dynamic gestures.

To improve robustness, landmark smoothing and temporal filtering techniques are applied to reduce jitter caused by rapid hand movements or partial occlusions. This ensures stable and consistent landmark trajectories, even under challenging conditions such as fast motion or imperfect lighting. The extracted landmark sequences serve as the primary input features for the deep learning model, enabling efficient and accurate gesture classification.

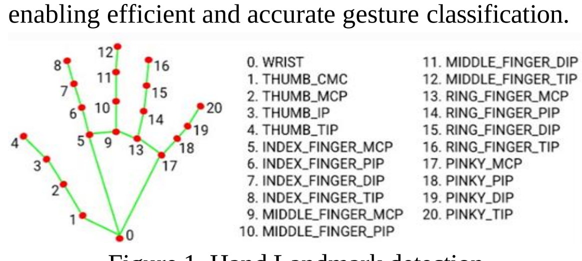

Figure 1. Hand Landmark detection

4.5. GESTURE CLASSIFICATION USING DEEP LERANING MODEL

The deep learning model architecture forms the core intelligence of the proposed Kannada Sign Language

(KSL) recognition system. It is designed to efficiently learn discriminative patterns from hand landmark features while maintaining real-time performance and scalability. The architecture focuses on capturing both spatial relationships among hand joints and, when extended, temporal dependencies across gesture sequences, enabling robust recognition under real-world conditions.

The model receives input in the form of normalized hand landmark coordinates extracted using MediaPipe. Each hand is represented by 21 key landmarks, with each landmark consisting of (x, y, z) coordinates. These coordinates are normalized with respect to the wrist joint to achieve translation and scale invariance. For static gesture recognition, a single frame's landmark set is used, resulting in a compact feature vector. For dynamic gestures, landmark coordinates from multiple consecutive frames are stacked to form a temporal sequence.

This landmark-based representation significantly reduces input dimensionality compared to raw image data, enabling faster training and inference while minimizing sensitivity to background noise and lighting variations. For static gestures, a Convolutional Neural Network (CNN) is employed to extract spatial features from given landmark data. The CNN architecture consists of multiple convolutional layers with small kernel sizes, which learn local geometric relationships between neighboring joints, such as finger curvature, inter-finger distances, and joint alignments.

Each convolutional layer is followed by Batch Normalization, to stabilize learning and improve convergence speed ReLU activation, to introduce non-linearity and enable the model to learn complex gesture patterns. Max Pooling, to reduce dimensionality and retain the most salient spatial features. This hierarchical feature learning allows the model to capture low-level joint relationships in early layers and more abstract hand configurations in deeper layers.

To support dynamic gestures and continuous sign recognition, the architecture can be extended by incorporating temporal modeling layers. Landmark sequences across frames are passed to recurrent neural networks such as Long Short-Term Memory (LSTM) or Gated Recurrent Units (GRU). These recurrent layers are capable of learning motion patterns, gesture transitions, and temporal dependencies that cannot be captured by static CNNs alone. By analyzing how hand landmarks evolve over time, the model can differentiate between gestures with similar hand shapes but different motion trajectories. Temporal dropout and sequence padding are applied to handle variable-length gesture sequences and reduce overfitting.

In scenarios where both spatial and temporal features are used, a feature fusion mechanism combines outputs from CNN and LSTM/GRU layers. This fusion allows the model to jointly consider hand posture and movement dynamics, leading to improved recognition accuracy, especially for complex gestures. Feature fusion is implemented using concatenation followed by dense layers that learn optimal feature weighting. This modular design enables easy expansion to multimodal inputs, such as facial landmarks or body pose features, in future system upgrades.

The proposed Kannada Sign Language recognition model is trained using a supervised learning approach, where each gesture sample is associated with a labeled class representing a specific Kannada sign. The training process employs the categorical cross-entropy loss function to quantify the difference between predicted gesture probabilities and ground-truth labels. This loss function is particularly well-suited for multi-class classification problems and enables effective gradient- based learning. Optimization is performed using adaptive optimization algorithms such as Adam, which dynamically adjust learning rates for individual parameters, resulting in faster convergence and improved training stability.

To ensure robust model performance and generalization, the dataset is systematically divided into training, validation, and testing subsets. The training set is used to learn model parameters, while the validation set is continuously monitored to assess model behaviour during training. Key performance indicators such as the validation loss and accuracy are tracked to detect overfitting or underfitting. Early stopping mechanisms are employed to automatically halt training when validation performance ceases to improve, thereby preventing unnecessary training and reducing the risk of overfitting. In cases where class imbalance exists due to uneven distribution of gesture samples, weighted loss functions and data augmentation strategies are applied to ensure fair learning across all gesture classes and to improve model robustness.

For real-time applicability, the model architecture is optimized to achieve low latency and efficient memory usage without compromising recognition accuracy. The use of hand landmark-based input significantly reduces computational complexity compared to pixel-based image models, enabling faster processing and making the system suitable for real-time execution on standard computing devices. This compact feature representation allows the model to perform inference rapidly while remaining resilient to environmental variations such as lighting and background noise.

To further enhance deployment efficiency, model optimization techniques such as pruning and quantization can be applied during the deployment phase. These techniques reduce model size and computational overhead, enabling smooth execution on mobile devices, embedded systems, or low-power hardware platforms. Once trained and optimized, the model is serialized and seamlessly integrated into the real-time recognition pipeline. This integration enables continuous inference from live camera input with minimal delay, providing immediate gesture-to-text translation and ensuring a responsive and user-friendly interaction experience.

V. PROPOSED METHODOLOGY

The proposed methodology presents a structured and efficient approach for developing a real-time Kannada Sign Language (KSL) recognition system using computer vision and deep learning techniques. The system is designed to accurately recognize hand gestures and translate them into corresponding Kannada text, while ensuring robustness, scalability, and real-time performance. The overall methodology consists of multiple sequential stages, including data acquisition, pre-processing, hand landmark detection, feature extraction, deep learning-based classification, and real- time output generation.

The methodology begins with data acquisition, where hand gesture images and video sequences corresponding to Kannada Sign Language alphabets and commonly used words are captured using a standard RGB camera. The dataset is collected from multiple users under varying lighting conditions, backgrounds, and hand orientations to ensure diversity and improve model generalization. Each gesture sample is carefully labeled to create a reliable ground truth for supervised learning.

Once the data is collected, pre-processing is performed to enhance data quality and consistency. This stage will include resizing and normalizing input frames, reducing noise, and applying data augmentation techniques such as rotation, scaling, and brightness adjustment. These steps help mitigate the effects of environmental variations and increase the robustness of the model. For dynamic gestures, temporal smoothing and frame alignment techniques are applied to maintain consistency across gesture sequences.

In the next stage, hand landmark detection is carried out using the MediaPipe framework. MediaPipe detects the hand region in each video frame and extracts 21 key hand landmarks representing finger joints and palm structure. This landmark-based representation focuses on essential hand features while eliminating background distractions. The extracted landmarks are normalized relative to a reference joint to ensure scale and translation invariance, making the system robust to changes in camera distance and hand position.

Following landmark detection, feature extraction and representation are performed. The spatial relationships between hand joints, such as distances and relative positions, are encoded into feature vectors. For static gestures, features from a single frame are sufficient, whereas dynamic gestures use sequences of landmark features across multiple frames to capture motion patterns and temporal dependencies.

The deep learning-based classification stage forms the core of the recognition system. A Convolutional Neural Network (CNN) is trained to learn discriminative spatial patterns from landmark features for static gesture recognition. For dynamic gestures, temporal modeling techniques are incorporated to analyze landmark sequences and recognize continuous hand movements. The model is trained using supervised learning, optimized using adaptive gradient-based methods, and evaluated using appropriate performance metrics to ensure accuracy and reliability.

Finally, the trained model is integrated into a real-time recognition and output generation module. During live operation, the system captures video input, processes frames in real time, detects hand landmarks, and classifies gestures using the trained model. The recognized gesture is mapped to its corresponding Kannada text and displayed instantly through a user-friendly interface. To improve stability and usability,

confidence thresholding and temporal smoothing techniques are applied to minimize prediction fluctuations.
Overall, the proposed methodology provides a robust, efficient, and scalable solution for Kannada Sign Language recognition. By combining landmark-based feature extraction, deep learning classification, and real- time processing, the system effectively addresses challenges related to environmental variability, computational efficiency, and user diversity. The modular design of the methodology also supports future enhancements such as dynamic sentence-level recognition, text-to-speech conversion, and multimodal integration, where making it a practical assistive communication tool for real-world applications.

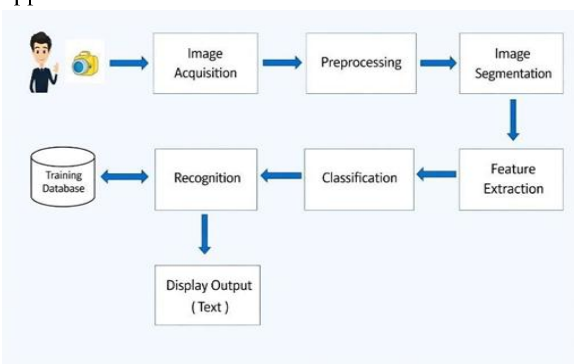

Figure 2. Proposed Methodology

VI. APPLICATIONS AND FUTURE SCOPE

The proposed Kannada Sign Language (KSL) recognition system has significant practical applications in improving communication accessibility for hearing- and speech-impaired individuals. By converting hand gestures into readable Kannada text in real time, the system helps bridge the communication gap between sign language users and the general population, enabling more inclusive day-to-day interactions without the constant need for human interpreters.
In educational environments such as schools, colleges, and special education centers, the system can serve as an effective assistive tool for both teachers and students. It can support classroom communication, help students learn Kannada sign language more efficiently, and enable educators who are not familiar with sign language to interact effectively with hearing-impaired learners.
The system can also be applied in healthcare settings to facilitate communication between medical professionals and hearing-impaired patients. Accurate interpretation of sign language gestures can assist in understanding patient concerns, providing medical instructions, and improving overall healthcare service quality. Similarly, the system can be deployed in public service areas such as banks, government offices, railway stations, and service centers to promote independent access to essential services for hearing-impaired individuals.
Additionally, the project has applications in human–computer interaction, where hand gestures can be used as an intuitive and contactless mode of interaction with digital systems. Integration with mobile and web-based platforms further enhances portability and usability, allowing users to access the system through smartphones or web applications. The system can also be used as a training and awareness tool to help non-signers learn Kannada sign language and promote inclusivity in society.
Although the proposed system effectively recognizes static Kannada sign language gestures, there is considerable scope for future enhancement. One major extension is the recognition of dynamic gestures and continuous sign sequences, which would enable sentence-level translation and more natural communication. Incorporating temporal models such as LSTM, GRU, or Transformer-based architectures can improve the handling of motion and gesture continuity.
Future work can also integrate text-to-speech technology to convert recognized Kannada text into spoken output, allowing hearing-impaired users to communicate verbally with non-signers. The system can be further expanded to support multiple regional or Indian sign languages, making it a multilingual sign language recognition platform suitable for a wider audience.

Optimizing the model for mobile and edge devices is another important direction, enabling offline processing and low-latency performance on smartphones and embedded systems. Expanding and standardizing the Kannada sign language dataset will further improve model robustness and generalization across different users, lighting conditions, and backgrounds. Moreover, integrating the system with IoT and smart environments can allow gesture-based control of devices, enhancing independence for hearing-impaired individuals. The adoption of advanced deep learning techniques and real- time feedback mechanisms can further improve accuracy, usability, and reliability, making the system more practical for real-world deployment.

VII. CONCLUSION

This project successfully presents the design and implementation of an automated Kannada Sign Language (KSL) recognition system using deep learning and computer vision techniques. The proposed system addresses the communication challenges faced by hearing- and speech-impaired individuals by enabling real-time recognition and translation of Kannada hand gestures into readable text. By leveraging MediaPipe for accurate hand landmark extraction and Convolutional Neural Networks for gesture classification, the system achieves reliable recognition of static Kannada sign language gestures with improved accuracy and efficiency.

The structured methodology involving data collection, pre-processing, feature extraction, model training, and real-time deployment ensures robustness and scalability of the system. The use of landmark-based feature extraction reduces background dependency and computational complexity, making the system suitable for real-time applications. Experimental results demonstrate that the model performs effectively under varying conditions, highlighting its potential for practical deployment in assistive communication scenarios.

Overall, the proposed system contributes to enhancing accessibility and inclusivity by providing a low-cost, camera-based solution for the Kannada Sign Language recognition. While the current implementation focuses on static gestures, the project establishes a strong foundation for future extensions such as dynamic gesture recognition, sentence-level translation, and speech synthesis. The outcomes of this work demonstrate the feasibility of applying deep learning techniques to regional sign language recognition and offer meaningful support toward bridging the communication gap between the hearing-impaired community and society.

REFERENCES

[1] D. Saraswathi, M. Anushree, H. R. Dishanth, S. Naidhile, and N. Thushar, “Kannada Sign Language Recognition Using Machine Learning,” 2025.

[2] H. C. Shruthi, J. M. Madhu, L. Adarsh, S. H. M. Shreeraksha, and T. N. Nagaratna, “Kannada Sign Language Recognition Using Hand Gestures,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 11, no. 4, pp. 1273–1278, 2024.

[3] R. Chaturvedi and R. Ramaiya, “Real-Time Indian Sign Language Recognition Using YOLOv11 and Keypoint Detection,” *IEEE Access*, 2025.

[4] S. Hon, M. Sidhu, S. Marathe, and T. A. Rane, “Real-Time Indian Sign Language Recognition Using Convolutional Neural Networks,” in *Proc. Int. Conf. on Intelligent Systems and Applications*, 2024.

[5] S. Biswas, A. Nandy, A. K. Naskar, and R. Saw, “MediaPipe and LSTM-Based Architecture for Real-Time Hand Gesture Recognition,” in *Proc. IEEE Int. Conf. on Signal Processing and Communication*, 2023.

[6] C. Hemanth, M. Jyothi, L. Adarsh, H. M. Shashikala, and N. Thejaswini, “Kannada Sign Language Recognition Using Hand Gestures,” *IRJET*, 2024.

[7] B. Subramanian, B. Olimov, S. M. Naik, S. Kim, K.-H. Park, and J. Kim, “An Integrated MediaPipe-Optimized GRU Model for Indian Sign Language Recognition,” *IEEE Access*, vol. 10, pp. 102345–102357, 2022.

[8] R. Katoch, P. Sharma, and A. Gupta, “Indian Sign Language Alphabet and Number Recognition Using SURF, BOVW, and SVM,” *Procedia Computer Science*, vol. 198, pp. 398–405, 2022.

[9] A. Puranik, A. Gawande, M. Gujarathi, K. Patani, and S. Rane, “Video-Based Sign Language Recognition Using Recurrent Neural Networks,” in *Proc. Int. Conf. on Communication and Signal Processing*, 2022.

[10] P. Nedungadi, R. Menon, and S. Raman, “Vision Transformer–Based Conversational Agent for Indian Sign Language Recognition,” *IEEE Transactions on Human–Machine Systems*, vol. 55, no. 1, pp. 34–45, 2025.

[11] E. L. Hmar, B. Gogoi, and N. R. Varte, "A Comprehensive Survey on Sign Language Recognition: Advances, Techniques and Applications," *International Journal of Engineering Research & Technology (IJERT)*, vol. 14, no. 8, Aug. 2025.

[12] S. Patra, A. Maitra, M. Tiwari, K. Kumaran, S. Prabhu, S. Punyeshwarananda, and S. Samanta, "Hierarchical Windowed Graph Attention Network and a Large-Scale Dataset for Isolated Indian Sign Language Recognition," 2024.

[13] U. Rastogi, R. P. Mahapatra, and S. Kumar, "Advanced Gesture Recognition in Indian Sign Language Using a Synergistic Combination of YOLOv10 with Swin Transformer Model," *Scientific Reports*, vol. 15, Art. no. 32894, 2025.

[14] P. P. Waghmare, A. M. Deshpande, S. Dubewar, and T. Dhaybar, "Deep Learning Approach for Combined Indian Sign Language Recognition and Video Generation Model," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 4, 2024.

[15] H. K. Vashisth, T. Tarafder, R. Aziz, M. Arora, and Alpana, "Hand Gesture Recognition in Indian Sign Language Using Deep Learning," *Engineering Proceedings*, vol. 59, no. 1, 2023.

[16] D. G., Ramar, A. Karpagam, S. J., S. R., and P. S., "Sign Language Recognition Using Convolutional Neural Network," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 17s, 2024.

[17] S. N. Swapna, P. Shivani, D. M. Karthik, and G. Shivateja, "An Effective Real Time Sign Language Recognition Using YOLO Algorithm," in *Proc. 3rd Int. Conf. on Optimization Techniques in the Field of Engineering (ICOFE)*, 2025.

[18] N. Swapna, P. Shivani, D. M. Karthik, and G. Shivateja, "An Effective Real Time Sign Language Recognition Using YOLO Algorithm," *SSRN Preprint*, Nov. 2024.

[19] "A Review on Advances in Indian Sign Language Recognition: Techniques, Models, and Applications," *IJRASET*, 2024.

[20] R. D. Karthika and L. Ashok Kumar, "Indian Sign Language Recognition Using Deep Learning Techniques," *International Journal of Computer Communication and Informatics*, 2025.

[21] D. Dhargalkar *et al*., "A Survey on Sign Language Recognition," *IJARCCE*, 2021.

[22] M. Oudah, A. Al-Naji, and J. Chahl, "Hand Gesture Recognition Based on Computer Vision: A Review of Techniques," *Journal of Imaging*, vol. 6, no. 8, 2020.

[23] A. Carbo and E. Nalisnick, "Improving Handshape Representations for Sign Language Processing: A Graph Neural Network Approach," 2025.

[24] "Real-Time Indian Sign Language Recognition Using CNNs for Communication Accessibility," *International Journal of Multidisciplinary Research & Analysis*, vol. 7, no. 9, Sept. 2024.

[25] N. Swapna and P. Shivani, "An Effective Real Time Sign Language Recognition Using YOLO Algorithm," *SSRN Preprint*, Nov. 15, 2024.

[26] G. Hugar, R. M. Kagalkar, and A. Das, "Comparative Study of Hybrid Deep Learning Models for Kannada Sign Language Recognition," *International Journal of Computational Intelligence Systems*, 2025.