

**FORM 2**

THE PATENTS ACT, 1970

(39 of 1970)

&

The Patent Rules, 2003

**COMPLETE SPECIFICATION**

(See section 10 and rule 13)

**TITLE OF THE INVENTION**

**“Real-Time Sign Language Translation System: Integrating Bi-LSTM, CNN with Large Language Models”**

**Applicant(s)**

<b>NAME</b>	<b>NATIONALITY</b>	<b>ADDRESS</b>
1. Ms. Diya Bhatia	Indian	Final Year Student, Department of CSE AIML, KIET Group of Institutions, Delhi- NCR, Ghaziabad, Uttar Pradesh, India-201206
2. Ms. Vanshika Goyal	Indian	Department of CSE AIML, KIET Group of Institutions, Delhi- NCR, Ghaziabad, Uttar Pradesh, India-201206
3. Ms. Akanksha Chaudhary	Indian	Department of CSE AIML, KIET Group of Institutions, Delhi- NCR, Ghaziabad, Uttar Pradesh, India-201206
4. Mr. Bikki Kumar	Indian	Department of CSE AI, KIET Group of Institutions, Delhi- NCR, Ghaziabad, Uttar Pradesh, India-201206

5. Mr. Rajeev Kumar Singh	Indian	Department of CSE AIML, KIET Group of Institutions, Delhi- NCR, Ghaziabad, Uttar Pradesh, India- 201206
---------------------------	--------	---

The following specification particularly describes the nature of the invention and the manner in which it is performed:

## **FIELD OF THE INVENTION**

**[0001]** The present invention is related to Indian Sign Language Translation in the Computer Science and Artificial Intelligence field.

## **BACKGROUND OF THE INVENTION**

**[0002]** Sign language becomes the most important one by which one can interact easily and foster accessibility and inclusive communication and which in turn, makes the deaf and hearing-impaired community experience the same communication and mutual understanding like the rest of the world.

**[0003]** The main focus in the first phase of the project is to develop an Indian Sign Language dataset (ISL) using the MediaPipe library. The data set is comprised of a fluent and diverse vocabulary that the kids are supposed to use to construct good meaningful sentences. The subsequent processes of data cleaning, normalization, and alignment keep the data accuracy which is essential for maintaining data consistency during machine translation and for providing good input for the next stage.

**[0004]** Bidirectional Long Short-Term Memory (Bi-LSTM) network and a Convolutional Neural Network (CNN) network is proposed. The integration of the Bi-LSTM gives the ability to capture long-range dependencies in a sequential dataset and the CNN infers the possibility of providing efficient image processing by combining other edge processes to be combined.

**[0005]** Few traditional methods are used to solve the problem. Also, AI-based methods are available for this problem, but the solutions are not up to the mark.

Every method has its challenges and nowadays no such solution exists that replaces human intervention in detecting the defect in cells and solar panels in the manufacturing industry.

**[0006]** Annotation has included few-shot prompting, which allowed the LLM to be trained to recognize the finer details of Indian Sign Language (ISL) and then to give contextually meaningful and grammatically correct translations. This step makes sure that the output is true semantically-right and the original gestures remain similar along with the correct grammar in English.

**[0007]** Sign language translation is an application that is both strong and can grow fast, hence it can handle several significant issues like the dataset is not enough, the gesture-to-word mapping and the lack of grammar between sign languages and natural languages. Through the integration of advanced deep learning techniques with state-of-the-art language models, this research is making a real meaning of the possibility of transformation in real-time applications of sign language translation.

**[0008]** Sign Language Recognition is a very active research area for many decades, starting from sensor-based approaches to computer vision and deep learning methods. The current methods are classified into Vision-Based and Non-Vision-Based approaches.

**[0009]** The arrival of Deep Learning, especially CNNs, changed everything. Garcia & Viesca (2016) employed fine-tuned GoogLeNet for American Sign Language (ASL) letter recognition with 98% accuracy on a small subset. Shagun

Katoch et al. (2022) employed Bag of Visual Words (BOVW) and SURF (Speeded Up Robust Features) for Indian Sign Language (ISL) recognition but only up to alphabets and digits, not words or sentences.

**[0010]** Most approaches word-for-word translate sign language, regardless of the unique syntax of sign languages like Indian Sign Language (ISL), which is different from spoken English grammatically.

## **OBJECTIVE OF THE INVENTION**

**[0011]** Most current models are oriented towards identifying isolated words, alphabets, or numbers but do not attempt to build full sentences in grammatically correct order.

**[0012].** Most approaches word-for-word translate sign language, regardless of the unique syntax of sign languages like Indian Sign Language (ISL), which is different from spoken English grammatically.

## **BRIEF DESCRIPTION OF DRAWINGS**

**[0013]** The system uses a unique hybrid CNN-LSTM architecture that combines several bidirectional LSTM layers improved by batch normalization with Conv1D layers for initial feature extraction. With an accuracy of 99.69% over 26 Indian Sign Language classes, this architecture is especially made for real-time sign language processing. The model maintains effective memory consumption appropriate for edge device deployment while incorporating optimized dense layers for classification and key spatial dropout layers for regularization.

## **DETAILED DESCRIPTION OF THE INVENTION**

The innovation is Sign Language Translation System to fill the gap between Indian Sign Language (ISL) speakers and non-signers. The technology uses a combination of deep

learning (CNNs, Bi-LSTMs), Generative AI (Large Language Models), and speech processing (gTTS, Speech Recognition) for real-time translation from/into ISL signs to/from English speech/text.

### System Overview

The model was built using TensorFlow's Sequential API, and it comprised:

CNN Input Layer to acquire spatial features of sign images.

Bi-LSTM Layers (128, 128, 64 neurons) to capture temporal dependencies of sign sequences.

Dropout Layers and Batch Normalization Layers to prevent overfitting and control learning (40% and 30% dropout rates).

L2 Regularization to enhance generalization by making the model insensitive to minor signing changes.

Adam Optimizer for efficient training, particularly for categorical classification.

## **We claim(s)**

- 1. Sign Language Detection Architecture:** The system uses a unique hybrid CNN-LSTM architecture that combines several bidirectional LSTM layers improved by batch normalization with Conv1D layers for initial feature extraction. With an accuracy of 99.69% over 26 Indian Sign Language classes, this architecture is especially made for real-time sign language processing. The model maintains effective memory consumption appropriate for edge device deployment while incorporating optimized dense layers for classification and key spatial dropout layers for regularization.
- 2. Feature extraction and landmark normalization:** Our novel approach to landmark processing presents a special body reference point normalization method that is tailored for a 195-point feature set (69 pose + 63 left hand + 63 right hand points). In order to increase processing speed, the algorithm purposefully removes face landmarks. The normalization process dynamically adapts to different body proportions using shoulder width as a reference, incorporating temporal smoothing for stable sequence processing and robust handling of varying lighting conditions.
- 3. Training Methodology:** A thorough 5-fold cross-validation technique is used in the training process, along with specialized learning rate scheduling that adjusts for performance plateaus. In order to provide optimal model convergence, this methodology integrates early stopping

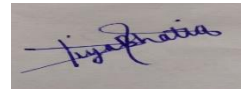
methods with best weight restoration. The system employs a balanced mini- batch sampling method, a bespoke loss function optimization, and strategic batch normalization for training stability and dropout regularization to avoid overfitting.

4. **System for Real-time Processing:** Our system uses optimal real-time normalization and effective Media Pipe landmark extraction to process videos continuously at 30 frames per second. Specifically built for mobile devices, the low-latency inference pipeline incorporates adaptive frame processing according to the capability of the device.
5. **Model Optimization:** Our optimization approach consists of a thorough architectural search for the best layer configuration and a methodical grid search for hyperparameter optimization. This includes careful balancing of performance-accuracy trade-offs and memory footprint reduction while maintaining real-time processing capabilities.
6. **Inference System:** The inference pipeline implements real-time video processing with efficient feature extraction and normalization, optimized for minimal latency. It incorporates sophisticated confidence score calculation and thresholding mechanisms, along with result post-processing and smoothing. The system includes comprehensive error detection and correction capabilities, supported by continuous performance monitoring and logging.



**Dated this 21<sup>st</sup> day of April 2025**

Signature:

A rectangular box containing a handwritten signature in blue ink. The signature appears to be 'Diya Bhatia' written in a cursive style.

**Applicant(s)**

Ms. Diya Bhatia et. al.

## **ABSTRACT**

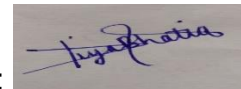
### **Real-Time Sign Language Translation System: Integrating Bi-LSTM, CNN with Large Language Models**

Sign language becomes the most important one by which one can interact easily and foster accessibility and inclusive communication and which in turn, makes the deaf and hearing-impaired community experience the same communication and mutual understanding like the rest of the world. However, despite incredible technical developments, there are still some issues with precise representation of the spatial-temporal features of this language and its translation into human language text. This research suggests a sign language interpretation mechanism bound together by a three-stage approach including data creation and preprocessing, model training, and sentence translation methodology.

The main focus in the first phase of the project is to develop an Indian Sign Language dataset (ISL) using the MediaPipe library. The data set is comprised of a fluent and diverse vocabulary that the kids are supposed to use to construct good meaningful sentences. The subsequent processes of data cleaning, normalization, and alignment keep the data accuracy which is essential for maintaining data consistency during machine translation and for providing good input for the next stage.

**Dated this 21<sup>st</sup> day of April 2025**

Signature:



**Applicant(s)**

Ms. Diya Bhatia et. al.