Intellectual Property Disclosure Form (Draft)

Title of the invention: Real-Time American Sign Language (ASL) Fingerspelling to Text using Two-Layer CNN with Frame-Consensus Logic

Innovator(s)

Norah Srivastava — B.Tech Student (Point of Contact)

Department: Electrical Engineering , IIT Kanpur

Address: Pune

Phone: 7058772004

Emails: norah22@iitk.ac.in

Course: DES646 (AI/ML for Designers)

Anwesha Dwivedi — B.Tech Student

Department: Mechanical Engineering, IIT Kanpur

Address: Mumbai

Phone: 8369159322

Emails: anweshad22@iitk.ac.in

Course: DES646 (AI/ML for Designers)

Anukalp Rai — B.Tech Student

Department: Electrical Engineering, IIT Kanpur

Address: Jhansi

Phone: 9044401409

Emails: ranukalp22@iitk.ac.in

Course: DES646 (AI/ML for Designers)

Bhavya Sharma — B.Tech Student

Department: Civil Engineering, IIT Kanpur

Address: Jaipur

Phone: 9730040742

Emails: bhavyas22@iitk.ac.in

Course: DES646 (AI/ML for Designers)

Ishan Mukundhan — B.Tech Student

Department: Civil Engineering, IIT Kanpur

Address: Delhi

Phone: 9822247635

Emails: ishanm22@iitk.ac.in

Course: DES646 (AI/ML for Designers)

Non-Confidential description (layman's language)

A webcam-based tool that reads hand signs (A–Z fingerspelling) in real time and converts them into on-screen text—without gloves or special sensors. A deep-learning model recognizes each letter, a temporal "frame-consensus" rule prevents flicker, and a "blank" gesture adds spaces so users can form words and sentences naturally.

Abstract (≤100 words)

This invention is a real-time ASL fingerspelling-to-text system using a two-layer convolutional architecture and temporal frame-consensus. A primary CNN classifies 26 letters (+ blank) from webcam frames; low-margin or confusable predictions (e.g., S/M/N; D/R/U) are routed to specialist models for disambiguation. A consensus gate confirms letters only after stable detection across N frames, while a blank gesture triggers spaces for word formation. Implemented with OpenCV and TensorFlow/Keras (optional autocorrect), the camera-only design targets robust, accessible HCI on consumer hardware.

Use Case

- Silent text entry for deaf/Hard-of-Hearing users in classrooms, clinics, and service counters.
- Real-time assistance for mixed ASL/non-ASL conversations.
- Low-cost HCI input modality for kiosks and public-service desks.
- Developer API for accessibility apps and educational tools.

Keywords

Webcam sign recognition; ASL fingerspelling; CNN; two-layer classifier; frame consensus; real-time HCI; OpenCV; TensorFlow; autocorrect.

Technical Description (for evaluation)

Relation to process/machine/composition:

A process for vision-based gesture acquisition, preprocessing, multi-stage classification, temporal consensus, and text formation; realized in software and deployable on commodity computers.

Detailed description:

1. Acquisition & Preprocess: 128×128 grayscale, Gaussian blur, ROI extraction.
2. Primary CNN: 27 classes (A–Z + blank), softmax probabilities.
3. Two-Layer Routing: low top-1 margin or class in {S,M,N}/{D,R,U}/{I,T,K,D} routes to a specialist CNN.
4. Temporal Consensus: confirm a letter when detected consistently for N frames (e.g., 50) with margin Δ.
5. Sentence Formation: sustained blank ⇒ space; optional autocorrect for word-level corrections.
6. Implementation Stack: Python, OpenCV, TensorFlow/Keras (+Hunspell optional).

Novelty:

- Two-layer disambiguation for visually similar ASL letters on consumer webcams.
- Frame-consensus gate that stabilizes outputs without extra sensors.
- Blank-driven spacing for natural word formation in real time.

Inventive Step (non-obviousness):

Combining specialist subset models with a confidence/margin-aware router and a temporal consensus gate tailored to ASL confusions enables robust accuracy without specialized hardware—beyond a standard single-CNN pipeline.

Advantages over comparable approaches:

- Camera-only; no glove/IMU—lower cost and complexity.
- Robust to confusable letters via targeted specialists.
- Temporal stability (reduced jitter) and built-in word spacing.
- Runs on consumer devices; simple deployment.

Patent search summary

Date: 09 Nov 2025
Databases: Google Patents, Espacenet, WIPO Patentscope
Queries (examples): "sign language recognition webcam", "ASL fingerspelling CNN", "gesture temporal consensus", CPC: G06V, G06K 9/, G10L 15/

Representative prior-art buckets & differentiation:

1. **Glove/IMU systems:** Wearable sensors capture hand pose. **Ours:** camera-only; no hardware.
2. **Single-CNN webcam letters:** One model classifies A–Z. **Ours:** two-layer design with specialist heads for confusable sets (e.g., D/R/U; S/M/N; I/T/K/D).
3. **Generic temporal smoothing (HMM/CRF/averaging):** Stabilize outputs post-hoc. **Ours:** margin-aware **frame-consensus gate** (commit on N frames + Δ margin) tuned for letter commits.
4. **Continuous sign→text translation:** Full dynamic signs to text. **Ours:** scoped to **fingerspelling** with **blank-driven spacing** and optional **autocorrect** for practical text entry.

Conclusion: Prior art covers sensors, single-stage classifiers, and generic smoothing. Our novelty is the **specific integration** of two-layer disambiguation + margin-aware frame consensus + blank-driven word formation on consumer webcams.

Experimental status / data

Working plan and architecture defined; dataset capture and model training described; real-time interface planned. (Prototype uploaded with code)

Technology Readiness Level (TRL)

TRL-3: Project Plan / Proof-of-Concept phase — device characteristics & proposal completed; POC in progress for prototype validation.

Need and Demand

- Bridges communication gap between ASL users and non-signers in education, healthcare, and services.

- Removes specialized-hardware barrier; suitable for low-resource settings and scalable deployments.
- Developer-friendly input channel for accessibility apps.

Market Access Information (high-level draft)

Global assistive-tech and accessibility-software markets are expanding; camera-only solutions have broader reach than sensor-based systems. Immediate addressable segments: education tech, public-service kiosks, telehealth/chat tools with accessibility features.

Future Developments

- Hand-shape segmentation/background subtraction for low-light/clutter.
- On-device optimization (TFLite/ONNX) for edge deployment.
- Expansion from fingerspelling to dynamic signs (temporal models).

Applications

Accessible text entry & live transcription; assistive HCI; education/training tools; customer-service counters.

IPR Ownership (disclosures)

Significant use of IITK funds/facilities? Course resources (standard lab machines/ webcams). No

Funding sources: NA

Inventor salary/remuneration: NA

Public disclosure: Mid-term proposal/presentation within course; no public publication yet.

Publication(s): None yet; considering academic paper (e.g., India HCI).

Sponsored/consultancy agreement with IITK? No.

Academic research toward a degree? Yes (course project component).

Revenue Sharing among Inventors (suggested)

- Norah Srivastava — 20%
- Anwesha Dwivedi — 20%
- Anukalp Rai — 20%
- Bhavya Sharma — 20%
- Ishan Mukundhan — 20%

Commercial Potential (brief)

Why would organizations procure this?

Low-cost, camera-only ASL text conversion that improves inclusivity at counters, classrooms, and apps.

Steps to commercialize:

- Package SDK/API; add calibration UI; harden models for diverse lighting.
- On-device optimization; pilot with an education/health partner.

Time to market:

~6–12 months post-POC with focused data collection & productization.

Prospective licensees / contacts:

None

PCT Filing

PCT/Foreign filing desired? No (at this time).

Declarations & Signatures

I/We agree to the IPDF terms and confirm no public disclosure beyond course submission/ presentation; IITK may act as applicant for filing and promotion/licensing as per policy.

Anukalp Rai

Bhavya Sharma

Anwesha Dwivedi

Norah Srivastava

Ishan R Mukundhan

9.11.2025