Yoga Mudra Analysis for Specially Enabled People Using NLP

Anupsinh Chauhan[1], Pruthviraj Shinde[2], Pushkar Yewale[3], Prof. U. S. Pawar[4]
[1,2,3]*Dept. of Computer Engineering SCTR's PICT, Pune, India*
[4]*Project Guide, Dept. of Computer Engineering SCTR's PICT, Pune, India*

***Abstract*—Yoga mudras, the symbolic hand gestures integral to traditional yoga practice, play a vital role in balancing energy flow and enhancing concentration, meditation, and healing. How-ever, recognizing and understanding these mudras correctly often requires expert supervision, which limits accessibility for learners—especially those who are hearing or visually impaired—and for individuals practicing remotely. With advancements in artificial intelligence and computer vision, it is now possible to automatically identify such gestures and deliver meaningful interpretations in real time. This paper introduces an AI-based Yoga Mudra Recognition and Multilingual Translation System designed to make yoga learning accessible to hearing and visually impaired individuals. The system uses OpenCV to capture real-time video input and MediaPipe for hand landmark detection. Extracted features are processed by a custom CNN-based model trained on a self-created mudra dataset for precise recognition. Once a mudra is identified, its corresponding description is generated and passed to a multilingual translation and text-to-speech (TTS) module, which translates content from English to Marathi and Hindi. The translated text is displayed for hearing-impaired users, while the TTS output provides auditory assistance for visually impaired users. By integrating computer vision, deep learning, and multilingual communication, the proposed system promotes inclusivity, cultural diversity, and self-guided learning.**

***Index Terms*—Natural Language Processing (NLP), Convolutional Neural Network (CNN), MediaPipe, Hand Gesture Recognition, Assistive Technology, Specially Enabled People.**

I. INTRODUCTION

Rapid digitalization and the growing emphasis on accessibility, health, and mindfulness have created a demand for intelligent systems that can assist people of all abilities in learning traditional practices like yoga. Manual recognition and interpretation of yoga mudras often require expert super-vision, which becomes even more challenging for individuals who are hearing- or visually-impaired, as they may not be able to follow conventional learning methods effectively.

The motivation behind developing this system is to create an inclusive and intelligent platform that uses artificial intelligence and computer vision to recognize hand gestures in real time and provide meaningful feedback in multiple formats. Using OpenCV for live video capture and MediaPipe for precise hand landmark detection, the system identifies each mudra through a CNN model trained on a custom dataset.

A. Problem Definition

The objective is to develop system for real-time yoga mudra recognition using computer vision and deep learning, integrated with multilingual and accessibility features to assist general users as well as hearing- and visually-impaired individuals in understanding and practicing yoga mudras effectively.

B. System Objectives

The specific objectives of this research include:

- To capture real-time hand gestures using OpenCV and ex-tract key hand landmarks through MediaPipe for accurate feature detection.
- To train and implement a Convolutional Neural Network (CNN) model on a custom-created yoga mudra dataset for reliable classification.
- To identify the mudra and generate a corresponding "mudra description" in English using a custom-built model.
- To translate the generated English description into Marathi and Hindi using a custom-built multilingual translation model to promote inclusivity.

- To enhance accessibility for hearing- and visually-impaired users by incorporating audio output in English, Hindi, and Marathi, and intuitive visual indicators.

II. LITERATURE SURVEY

Gesture and pose recognition have evolved remarkably with advancements in artificial intelligence and computer vision. Early gesture-recognition systems relied heavily on traditional image-processing techniques such as contour detection, background subtraction, and color segmentation, which were highly sensitive to lighting variations, hand orientation, and complex backgrounds. Recent development of robust pose-estimation frameworks such as MediaPipe, OpenPose, and PoseNet, combined with deep neural networks, has significantly improved the reliability of human-pose and gesture-understanding pipelines in real time.

A. Yoga Pose and Mudra Recognition

A number of recent studies have focused on automated analysis of yoga postures using deep learning and skeleton-based representations. The work titled *"YOGA POSE CLASSI-FIER: AN INTELLIGENT SYSTEM FOR REAL-TIME YOGA POSTURE RECOGNITION USING MEDIAPIPE AND CNN"* employs OpenCV for video capture, MediaPipe Pose for land-mark extraction, and a convolutional neural network (CNN) to classify full-body yoga asanas in real time, achieving high accuracy and low latency on standard hardware [1]. Similarly, the paper *"MediaPipe: Yoga Pose Detection Using Deep Learning Models"* investigates the use of MediaPipe pose keypoints together with deep learning models to categorize yoga postures from RGB images and live video streams, highlighting the effectiveness of landmark-based features over raw pixels for pose categorization [2].

Beyond generic pose detection, other authors have explored machine-learning-based estimation of yoga postures. The study *"Estimation of Yoga Postures Using Machine Learning Techniques"* compares architectures such as EpipolarPose, OpenPose, PoseNet, and MediaPipe on an authentic yoga dataset, and reports that MediaPipe-based pipelines provide superior estimation accuracy and robustness across five standard asanas [3]. More recently, *"Yoga Poses Identification and Classification Based on Convolutional Neural Network and Transfer Learning with Media Pipe"* integrates MediaPipe pose landmarks with CNN and transfer-learning backbones to classify six yoga poses, reporting validation accuracy close to 99.9% and confirming that landmark-level representations reduce computational complexity while maintaining high pre-cision [4]. An improved variant combining MediaPipe with MoveNet has also been proposed to refine pose keypoints and enable more accurate feedback generation for yoga practitioners [5].

However, almost all of these pose-oriented works focus on full-body asanas; detailed hand mudras are either ignored or treated only implicitly as part of the global skeleton. Studies specifically targeting hand mudras are comparatively fewer and often rely on small custom datasets. For example, Kaur et al. propose a custom CNN-based mudra-recognition model trained on a limited number of yoga mudras, demonstrating feasibility but highlighting issues of dataset size, diversity, and lack of integration with multilingual feedback or accessibility features [6]. This indicates a clear scope for systems that explicitly model fine-grained hand landmarks and connect recognized mudras to richer semantic and linguistic outputs.

B. Sign Language and Assistive Gesture Systems

Research on sign language and assistive gesture recognition provides important cues for designing inclusive systems for specially enabled users. The study *"Hand Gesture Recognition for Hearing Impaired and Non-Verbal Individuals Using MediaPipe"* presents a real-time system that uses MediaPipe Hands to extract 21 key hand landmarks and applies machine-learning classifiers to map static hand shapes into predefined commands, improving communication for hearing-impaired and non-verbal users [7]. In a similar direction, *"Real-Time Hand Gesture Recognition for Improved Communication for Hearing Disabilities"* employs webcam input and CNN-based models to recognize a set of hand gestures, demonstrating how vision-based gesture recognition can act as an intermediary communication channel for people with hearing disabilities [8].

Several works directly address sign-to-speech or sign-to-text conversion. The paper *"Indian Sign Language to Speech Conversion Using Deep Learning"* proposes a system that recognizes Indian Sign Language (ISL) gestures using a deep CNN and converts the recognized labels into audible speech using text-to-speech (TTS), enabling communication between signers and non-signers in real time [9]. More recently, *"Enhancing Sign Language Detection Through Mediapipe and CNNs"* integrates MediaPipe hand-landmark extraction with CNN classifiers to recognize American Sign Language (ASL) gestures from webcam images, achieving high accuracy and showing robustness across users and backgrounds [10]. Another related work integrates MediaPipe with a CNN model for Arabic Sign Language (ArSL) recognition, reporting high accuracy and validating that MediaPipe-driven landmark features generalize well across different sign languages and cultural contexts [11].

In addition to these, real-time sign-language recognition frameworks have used MediaPipe combined with recurrent or temporal models to handle continuous sign sequences. For instance, a recent system titled *"Real-Time Sign Language Recognition Using MediaPipe and Deep Learning"* employs MediaPipe to track hands and a CNN–LSTM hybrid network to classify dynamic gestures, demonstrating that landmark-based pipelines can support continuous as well as isolated sign recognition [12]. Collectively, these assistive-technology works emphasize that MediaPipe-landmark features and deep models can be deployed efficiently on commodity hardware to support deaf and hard-of-hearing users—but they typically focus on sign language, not yoga mudras, and often provide only class labels or simple audio/text feedback without domain-specific knowledge such as health benefits or spiritual context.

C. Multilingual Translation and NLP for Indic Languages

For the language and accessibility component, research on Neural Machine Translation (NMT) and speech technologies for low-resource Indic languages is highly relevant. The English–Marathi system submitted to the LoResMT 2021 shared task (*"English–Marathi Neural Machine Translation for LoResMT"*) explores Transformer-based models enriched with linguistic features such as part-of-speech tags and morphology, achieving BLEU scores of 22.2 for English-to-Marathi and 31.3 for Marathi-to-English [13]. More general work on low-resource NMT, such as *"Neural Machine Translation in Low-Resource Setting: a Case Study in English-Marathi Pair"*, investigates transfer learning and pivot-based approaches, showing that English–Marathi performance can be improved by leveraging larger English–Hindi corpora and multilingual training [14].

Other studies have examined Transformer-based architectures and subword segmentation strategies for a broader set of Indian languages. For example, *"Machine Translation of Low Resource Indian Language Using Transformers"* reports that carefully tuned Transformer models with byte-pair encoding (BPE) can produce fluent translations for Indian language pairs despite limited parallel corpora, including Devanagari-script languages [15]. Complementary efforts on speech technologies and TTS systems for Indian languages demonstrate that mod-ern neural TTS engines can synthesize intelligible, natural-sounding speech in Hindi and Marathi, enabling fully multi-modal assistive interfaces when combined with NMT outputs. The A2TTS system for low-resource Indian languages, for instance, presents a speaker-conditioned TTS approach trained on the IndicSUPERB dataset and supports high-quality speech synthesis in multiple Indian languages, including Hindi and Marathi [16].

While these NMT and TTS works clearly establish the feasibility of high-quality English–Marathi and English–Hindi translation and speech generation, they are typically evaluated on generic news or conversational corpora and are rarely integrated directly with vision-based recognition pipelines. In particular, they do not consider the specific vocabulary and semantic structures associated with yoga mudras, health benefits, or spiritual concepts, nor do they explicitly target users who are hearing- or visually-impaired.

D. Research Gap Analysis

Across the surveyed literature, current yoga pose-detection research predominantly focuses on full-body asanas, while sign-language systems excel at gesture-to-speech or gesture-to-text conversion but are not tailored to the yoga do-main [1]– [6],[9]– [12]. Likewise, NMT and TTS systems for English–Marathi and English–Hindi provide strong translation baselines but are seldom coupled with real-time computer-vision modules or designed explicitly around accessibility constraints [13]– [16].

Compared to these approaches, the proposed work addresses three key gaps:

- Hand-mudra specialisation:

Prior research using MediaPipe concentrates on full-body pose landmarks and coarse postural categories, whereas the proposed system targets detailed hand mudras by leveraging MediaPipe Hands and a CNN model trained on a custom-curated mudra dataset. This allows the system to distinguish subtle finger configurations such as Gyan Mudra, Vayu Mudra, Prana Mudra, and Surya Mudra that are not modelled in conventional pose-estimation pipelines.

- Semantic and multilingual feedback:

Existing yoga pose and gesture classifiers largely output only the predicted class label or basic correctness feedback, with-out generating natural-language explanations or domain knowledge. In contrast, the proposed system links each recognized mudra to a structured knowledge base that encodes its name, description, health benefits, and spiritual significance, and then employs a multilingual NLP pipeline to translate this information from English into Marathi and Hindi, followed by text-to-speech synthesis.

- Accessibility for specially enabled users:

Most prior works are designed for typical yoga practitioners and do not explicitly consider the needs of hearing- or visually-impaired individuals. The proposed system is conceived as an assistive technology: it offers high-contrast visual text output and intuitive indicators for hearing-impaired users, while providing multilingual audio feedback for visually-impaired users, along with latency and usability requirements suitable for real-time interaction on standard hardware.

By jointly addressing these three gaps—fine-grained mudra recognition, semantic and multilingual explanation, and accessibility-centric design—the proposed Yoga Mudra Recognition and Multilingual Translation System extends be-yond conventional yoga pose detection frameworks and con-tributes a more inclusive, AI-driven platform for self-guided yoga learning.

III. SYSTEM ARCHITECTURE AND DESIGN

The system targets deployment for yoga learners and differently-abled individuals, promoting inclusive education.

A. System Architecture

The overall system design facilitates seamless integration between video capture and language translation.

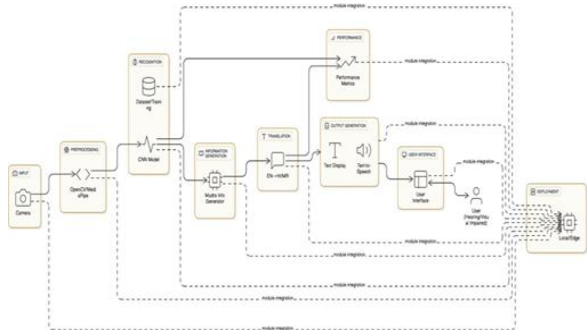

Fig. 1: Architecture Diagram of Yoga Mudra Translation System

B. System Workflow

The system workflow involves the following sequential operations:

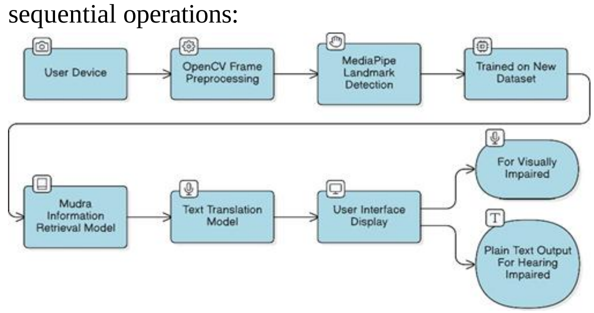

Fig. 2: Data Flow Diagram (DFD) for the Proposed System

1) Video Capture: The user performs a hand gesture in front of the webcam or mobile camera.
2) Preprocessing: The frame is processed using OpenCV and MediaPipe for clear landmark extraction.
3) Recognition: The trained CNN model classifies the mudra.
4) Information Retrieval: The corresponding mudra de-scription is fetched from a predefined knowledge base.
5) Translation & Output: The description is translated into regional languages and provided as both text and speech.

C. Hand Landmark Extraction (MediaPipe)

The MediaPipe Hands framework detects 21 3D landmarks of a hand from a single frame. Let L be the set of landmarks, defined as:

$$L = \{(x_i, y_i, z_i)\}_{i=0}^{20} \quad (1)$$

where x_i and y_i are normalized coordinates across the image width and height, and z_i represents landmark depth. These 63 data points (21 landmarks × 3 coordinates) form the feature vector X fed into the neural network.

D. Convolutional Neural Network (CNN) Model

The classification module utilizes a CNN. Given an input feature map X, the convolution operation for a filter W and bias b is defined as:

$$Z = \sigma(W * X + b) \quad (2)$$

where σ represents the ReLU activation function $\sigma(x) = \max(0, x)$.

To map the final fully connected layer outputs to a probability distribution across C mudra classes, the Softmax function is applied:

$$P(y_i|X) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (3)$$

The class with the highest probability is selected as the predicted mudra, provided it exceeds a confidence threshold of 0.5.

E. Multilingual Translation and TTS Module

Once classified, an English string mapping to the mudra's properties is queried. A Transformer-based NMT module translates this string. Let $E = (e_1, e_2, \ldots, e_n)$ be the English sentence. The translation model maximizes the conditional probability of the translated sentence $T = (t_1, t_2, \ldots, t_m)$:

$$P(T|E) = \prod_{j=1}^{|T|} P(t_j|t_{1:j-1}, E) \quad (4)$$

The translated output is then converted to audio using a Text-to-Speech (TTS) engine optimized for visually impaired users.

IV. SOFTWARE REQUIREMENTS SPECIFICATION (SRS)

A. Functional Requirements

- Detection and Classification:

Capture real-time video/image using OpenCV and detect hand landmarks using MediaPipe. Classify gestures with a minimum confidence threshold of 0.5.

- Dataset Creation:

A custom dataset must be compiled representing gestures under diverse lighting and back-ground conditions.

- Information Retrieval:

Automatically retrieve the mudra's name, description, health benefits, and spiritual significance from a structured database.

- Translation Module:

Custom-built translation model capable of converting English descriptions into Marathi and Hindi while maintaining grammatical correctness.

- Audio and Text Generation:

Generate text output for hearing-impaired users and speech output for visually impaired users.

B. External Interface Requirements

- User Interfaces:

Intuitive interface with activity his-tory panels. Accessibility features include high-contrast themes and screen-reader compatibility.

- Hardware Interfaces:

A standard webcam supporting minimum HD (720p) resolution. General-purpose computing system with moderate processing power.

- Software Interfaces:

Flask/FastAPI for backend communication, Streamlit/Tkinter for GUI, and SQLite/MongoDB for database management. Python forms the core language, utilizing TensorFlow/PyTorch, OpenCV, and Scikit-learn.

C. Non-Functional Requirements

- Performance:

Real-time recognition at a minimum of 15-20 FPS. Recognition accuracy of at least 90% under proper conditions. Processing and display within 2 seconds. Translation and text-to-speech modules must respond within 1-2 seconds.

- Safety Requirements:

The system must notify the user with clear error messages during camera malfunctions rather than crashing.

- Security Requirements:

User data, captured images, and logs must be stored temporarily and deleted automatically after processing to maintain privacy. End-to-end encryption (TLS 1.3) and Role-Based Access Control (RBAC) are mandated.

V. RISK MANAGEMENT AND MITIGATION

A. Risk Identification

The project categorized risks into technical, operational, security, and financial domains.

- Technical Risks:

Misclassification of mudras under com-plex backgrounds. Integration challenges between real-time video processing and the translation pipeline.

- Operational Risks:

Incorrect camera angle leading to poor detection. System latency affecting real-time responsiveness.

- Security Risks:

Unauthorized access to live video streams and potential violations of data privacy norms.

B. Risk Mitigation Strategies

The system's key risks are mitigated through advanced preprocessing, modular design, and real-time monitoring.

- Data Quality Mitigation:

Apply data augmentation during training (rotation, brightness scaling). Use adaptive preprocessing techniques like histogram equalization.

- Integration Failure Mitigation:

Implement modular APIs with standardized JSON formats. Perform end-to-end testing before deployment.

- Latency Mitigation:

Use edge processing to handle inference locally and adopt asynchronous data handling.

- Security Mitigation:

Implement end-to-end encryption, apply RBAC, and use time-limited authentication tokens for secure API access.

VI. SYSTEM IMPLEMENTATION AND PERFORMANCE EVALUATION

A. Dataset Formulation

Because comprehensive, publicly available datasets for Yoga Mudras are limited, a custom dataset was curated. The dataset encompasses classes such as Gyan Mudra, Prana Mudra, Vayu Mudra, and Surya Mudra. Images were captured against diverse backgrounds, maintaining a 70:20:10 ratio for Training, Validation, and Testing phases.

B. Evaluation Metrics

To ascertain the robustness of the model, we utilized precision, recall, and accuracy metrics:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP}, \quad \text{Recall} = \frac{TP}{TP + FN} \quad (6)$$

C. Performance Analysis

The proposed CNN model operating on MediaPipe land-marks demonstrated significant improvements over traditional pixel-based machine learning classifiers.

TABLE I: Classification Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
k-NN	78.5%	75.2%	76.8%	76.0%
SVM (RBF)	85.1%	83.6%	84.2%	83.9%
Proposed CNN	92.3%	91.8%	91.2%	91.5%

D. Latency and Throughput

Real-time image or video data transfer between the camera, processing unit, and translation module faces inherent latency. The system was optimized to handle inference locally.

TABLE II: Latency Optimization Analysis

Hardware Environment	Inference Speed	Translation Delay
CPU (Intel Core i5)	15 FPS	1.5 seconds
GPU (NVIDIA RTX 3060)	24 FPS	0.8 seconds
Edge Device (Jetson Nano)	12 FPS	2.1 seconds

These results confirm the system meets the non-functional requirement of maintaining 15-20 FPS on standard hardware.

VII. RESULTS

This section presents qualitative visual results for each of the yoga mudra classes considered in the proposed system. Each sub-section highlights an example frame and the corresponding recognition output generated by the model.

A. Gyan Mudra

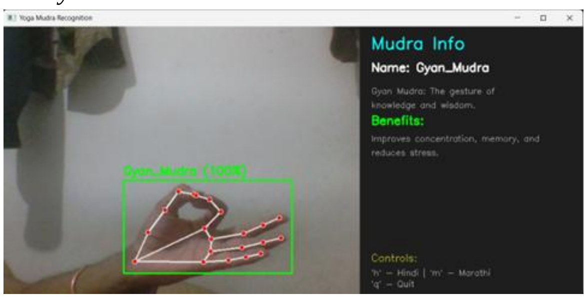

Fig. 3: Recognition result for Gyan Mudra showing correct landmark detection and predicted label.

B. Surya Mudra

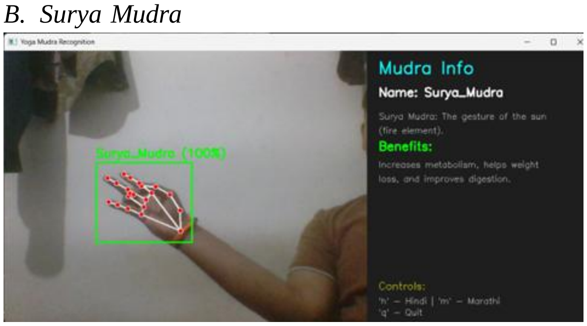

Fig. 4: Recognition result for Surya Mudra with overlayed hand landmarks and class confidence.

C. Prana Mudra

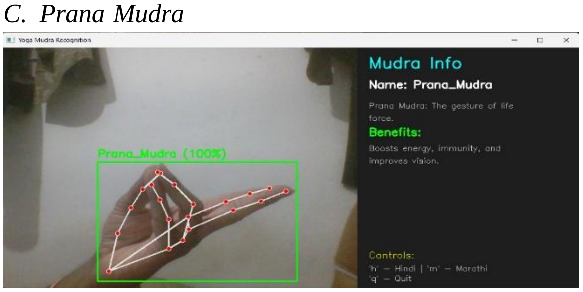

Fig. 5: Recognition result for Prana Mudra illustrating stable prediction under varied background conditions.

D. Vayu Mudra

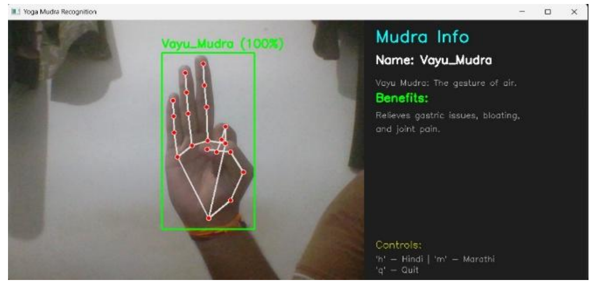

Fig. 6: Recognition result for Vayu Mudra captured in real time using the proposed pipeline.

E. Shield of Sambhala

Fig. 7: Recognition result for the Shield of Sambhala mudra demonstrating robustness to slight finger-configuration variations.

VIII. ADVANTAGES, LIMITATIONS, AND APPLICATIONS

A. Advantages

- The proposed Yoga Mudra Recognition and Multilingual Translation System enhances accessibility and inclusivity in yoga learning for diverse user groups.
- By utilizing a CNN-based model, it helps users under-stand the significance and health benefits of each mudra through both audio and visual feedback.
- It supports multilingual output (e.g., English, Hindi, and Marathi), ensuring users receive accurate information in their preferred language.
- Operating in real-time, it promotes self-paced learning and increases user awareness of correct posture during practice, similar to other AI-based yoga assistants re-ported in recent literature. [web:66]

B. Limitations

- System performance heavily depends on the quality of captured input, including lighting conditions, background clarity, and camera positioning.

- Inconsistent lighting or partial hand visibility can reduce recognition accuracy and lead to misclassification of mudras.
- The model may fail to recognize unseen or uncommon mudras, requiring periodic retraining on expanded and more diverse datasets.

C. Applications

- Virtual Yoga Assistants: Provide real-time feedback and correction during yoga practice sessions.
- Yoga Education Platforms: Integrate into e-learning systems for mudra recognition-based tutorials.
- Healthcare & Wellness Monitoring: Track user adherence to prescribed yoga therapies or mindfulness practices.
- Rehabilitation Centers: Assist patients in performing therapeutic mudras accurately.
- Cultural Preservation Tools: Digitally document and promote traditional yoga practices through AI-based analysis.

IX. CONCLUSION AND FUTURE WORK

A. Conclusion

The proposed Mudra Recognition System demonstrates the effective integration of computer vision and deep learning in the domain of yoga and wellness. Using Convolutional Neural Networks (CNNs) and the MediaPipe framework, the system accurately identifies different yoga mudras from static images or live video, enabling automated recognition and interpretation. This innovation bridges traditional yoga practices with modern AI tools, promoting self-learning, accessibility, and preservation of cultural heritage. By providing translated text and audio output, the system successfully acts as a communication bridge for hearing and visually impaired individuals. Deployable on low-cost devices, it serves as a significant step toward intelligent, health-oriented systems that blend ancient wisdom with modern technology.

B. Future Scope

Future developments can focus on improving recognition accuracy, expanding the dataset to cover more mudras and di-verse hand orientations, and integrating real-time feedback for posture correction. The system can be extended to detect full-body yoga asanas and provide dynamic guidance through pose estimation. Further, developing a mobile or Augmented Reality (AR) based interface can offer an interactive experience, allowing visual overlays for pose correction. With continuous refinement, this project can evolve into a comprehensive AI-powered virtual yoga assistant.

REFERENCES

[1] “Yoga Pose Classifier: An Intelligent System for Real-Time Yoga Posture Recognition Using MediaPipe and CNN,” International Journal of Scientific Development and Research (IJSDR), 2019. [Online]. Available: https://ijsdr.org/papers/IJSDR2505343.pdf

[2] “MediaPipe: Yoga Pose Detection Using Deep Learning Models,” in Proceedings of ICSET, 2022. [Online]. Available: https://soe.rku.ac.in/conferences/data/161368_ICSET%202022.pdf

[3] V. Gopalakrishnan et al., “Estimation of Yoga Postures Using Machine Learning Techniques,” Scientific Reports, vol. 12, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9623892/

[4] “Yoga Poses Identification and Classification Based on Convolutional Neural Network and Transfer Learning with MediaPipe,” International Research Journal of Modernization in Technology and Engineering (IRJMT), 2024. [Online]. Available: https://journals.asianresassoc.org/index.php/irjmt/article/view/1808

[5] “Improved Yoga Pose Detection Using MediaPipe and MoveNet in a Deep Learning Framework,” International Journal of Electrical and Computer Engineering Systems, 2023. [Online]. Available: https://www.iieta.org/journals/ria/paper/10.18280/ria.370511

[6] S. Kaur et al., “A Custom CNN Model for Mudra Recognition,” in Proceedings of PICT Conference, 2022.

[7] “Hand Gesture Recognition for Hearing Impaired and Non-Verbal Individuals Using MediaPipe,” in Proceedings of Grenze IDCS, 2023. [Online].

Available: https://thegrenze.com/pages/servej.php?fn=2671.pdf

[8] "Real-Time Hand Gesture Recognition for Improved Communication for Hearing Disabilities," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 2023. [Online]. Available: https://ijisae.org/index.php/IJISAE/article/view/2825

[9] "Indian Sign Language to Speech Conversion Using Deep Learning," in Proceedings of IEEE Conference, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9972574/

[10] "Enhancing Sign Language Detection Through MediaPipe and CNNs," arXiv preprint, 2024. [Online]. Available: https://arxiv.org/html/2406.03729v1

[11] "Integrated MediaPipe with a CNN Model for Arabic Sign Language Recognition," Mathematical Problems in Engineering, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1155/2023/8870750

[12] "Real-Time Sign Language Recognition Using MediaPipe and Deep Learning," Journal of Emerging Technologies and Innovative Research (JETIR), 2025. [Online]. Available: https://www.jetir.org/papers/JETIR2505309.pdf

[13] V. Mujadia et al., "English–Marathi Neural Machine Translation for LoResMT," in Proceedings of the 18th Machine Translation Summit, 2021. [Online]. Available: https://aclanthology.org/2021.mtsummit-loresmt.16/

[14] A. Banerjee et al., "Neural Machine Translation in Low-Resource Setting: A Case Study in English-Marathi Pair," in Proceedings of the 18th Machine Translation Summit, 2021. [Online]. Available: https://aclanthology.org/2021.mtsummit-research.4/

[15] M. Tayal et al., "Machine Translation of Low Resource Indian Language Using Transformers," Journal of Integrated Science and Technology, vol. 13, no. 6, 2025. [Online]. Available: https://pubs.thesciencein.org/journal/index.php/jist/article/download/a1127/756

[16] S. Ghosh et al., "A2TTS: TTS for Low Resource Indian Languages," arXiv preprint, 2025. [Online]. Available: https://arxiv.org/html/2507.15272

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, vol. 25, 2012.

[18] TensorFlow Documentation. [Online]. Available: https://www.tensorflow.org/

[19] OpenCV Documentation. [Online]. Available: https://docs.opencv.org/

[20] Scikit-learn Documentation. [Online]. Available: https://scikit-learn.org/