HEALING HANDS: YOGA MUDRA DETECTION AND CORRECTION USING AI/ML

Project Proposal Reference No: *49S_BE_0794*

College : New Horizon College of Engineering, Bengaluru
Branch : Department of Computer Engineering
Guide(s) : Dr. T Kavitha
Student(s) : Ms. Thoppe Sridhar Srivarshine
Ms. Nivedita Niran
Ms. Pranjali Sinha
Ms. Preeyanka Shree

Keywords:
Yoga Mudra Detection, Hand Gesture Recognition, MediaPipe, MobileNetV2, TensorFlow.js, CNN, Real-Time Feedback, Client-Side Inference, MERN Stack, Computer Vision.

Introduction:
Yoga mudras are precise hand gestures rooted in ancient Indian practice, believed to regulate energy flow within the body and promote mental and physical well-being. Despite their simplicity in appearance, correct execution demands exact finger articulation. With the growing shift toward digital wellness and remote learning, practitioners increasingly lack access to expert instructors who can provide real-time corrective guidance.

Existing systems for gesture recognition either focus solely on detection without feedback, or rely on computationally intensive architectures unsuitable for browser-based deployment. Static online resources such as videos and images only provide passive guidance, offering no interactivity. Prior work using models like YOLOv7, Faster R-CNN, and ResNet-50 demonstrated high accuracy but introduced latency and scalability concerns, limiting practical use on consumer-grade devices.

"Healing Hands" addresses these gaps by proposing a lightweight, browser-based yoga mudra detection and correction system. It performs client-side inference using TensorFlow.js integrated with MediaPipe for hand landmark detection, eliminating the need for server-side processing. The system recognizes 8 mudra classes in real time and provides immediate corrective feedback, making it accessible to users regardless of their hardware capability. The project also integrates a MERN stack backend for session persistence and user progress tracking.

Objectives of the project :

1. Develop a client-side, browser-based yoga mudra recognition system that requires no server-side processing, ensuring privacy and low latency.
2. Integrate MediaPipe hand landmark detection with a CNN classifier to accurately identify 8 yoga mudras in real time.
3. Implement a Geometric Heuristic Verification layer to disambiguate visually similar mudras using Euclidean distance metrics.
4. Provide real-time corrective feedback to guide users toward accurate mudra execution without the need for an instructor.
5. Build a MERN stack web application with session tracking to monitor user progress and improvement over time
6. Make yoga practice more accessible to individuals who lack access to physical trainers or classes.

Methodology:

1. System Architecture: The application is built using Next.js for the frontend. All inference runs client-side via TensorFlow.js, eliminating frame transmission to remote servers. MongoDB Atlas stores user authentication states and session progress via RESTful API endpoints.
2. Hand Tracking and Feature Extraction: MediaPipe extracts 21 3D hand landmarks per frame from the webcam's RGB input stream. A dynamic Region of Interest (ROI) bounding box is computed from detected landmarks with padding to capture finger extensions. The cropped region is resized to 224×224 pixels and normalized to [0,1] before being passed to the classifier.
3. CNN Classification: The primary classifier is a MobileNetV2-based CNN, optimized for web execution and trained on a custom dataset of ~1,000 images across 8 mudra classes (Chin, Jnana, Vayu, Prana, Prithvi, Surya, Varuna, Apana). The model outputs a softmax probability vector, with detections accepted above a confidence threshold of 0.5. The dataset was collected under varied lighting conditions and split 85:15 for training and testing. Augmentation techniques including random rotations (±15°), brightness adjustments, horizontal flipping, and zoom transformations were applied to improve robustness.
4. Geometric Heuristic Verification: A secondary validation layer uses Euclidean distance metrics between key landmarks to resolve misclassifications between similar mudras (e.g., Surya vs. Prithvi). A dynamic hand scaling factor (wrist-to-middle MCP distance) ensures scale invariance. Fold detection checks if a finger tip is closer to the wrist than its MCP joint, and touch detection checks if fingertip-to-thumb distance falls below 45% of the scaling factor.

5. Temporal Stabilization: A majority voting buffer over the last 10 frames (FIFO queue) filters detection noise. A 2-second grace interval prevents the practice timer from resetting on momentary detection loss, ensuring a smooth user experience.

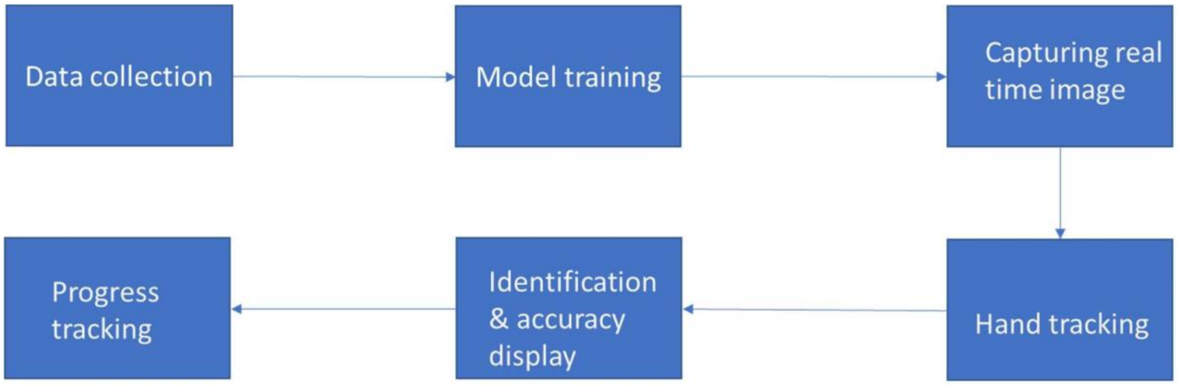

Figure.1: System Architecture

Results and Conclusions

The trained MobileNetV2-based CNN achieved approximately 92% classification accuracy across all 8 mudra classes, outperforming generic CNN and Teachable Machine baselines. Precision, recall, and F1-scores indicated minimal class imbalance. The confusion matrix confirmed strong separability between most mudras, with minor misclassifications occurring between Jnana and Chin mudras due to their high finger-placement similarity.

In real-time performance evaluation, the system maintained inference latency well within acceptable thresholds on mid-range consumer devices, with no noticeable frame rate degradation — validating the feasibility of client-side deployment. The proposed system demonstrated the smallest model size among compared architectures (Faster R-CNN, YOLOv7, Generic CNN), directly contributing to lower memory consumption and faster inference.

User improvement analysis showed that confidence scores increased with repeated practice sessions, and session logs indicated a reduction in classification uncertainty over time, supporting the effectiveness of the real-time corrective feedback mechanism.

In conclusion, Healing Hands successfully demonstrates that a lightweight, browser-based AI system can detect yoga mudras accurately and provide real-time corrective feedback without high computational requirements, making it accessible, scalable, and deployable across multiple platforms.

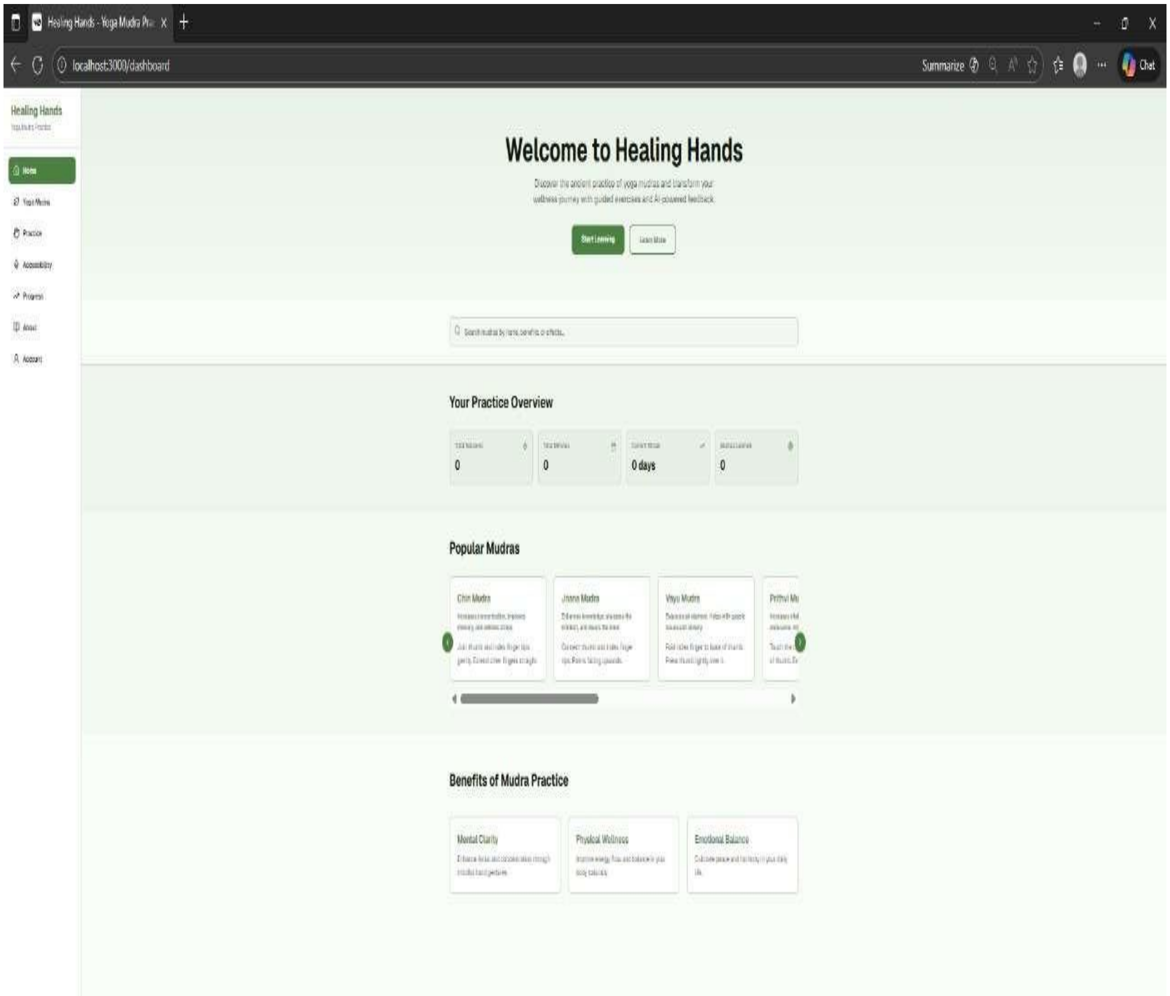

Figure.2: Home Page

Project Outcome & Industry Relevance:

Healing Hands demonstrates that production-ready AI can be deployed entirely within a browser without sacrificing real-time performance, which has meaningful implications for the digital health and EdTech industries. The system bridges a genuine accessibility gap — millions of yoga practitioners globally lack access to certified instructors, and this platform provides an always-available, intelligent substitute for basic mudra guidance.

From a research standpoint, the project contributes a novel hybrid architecture combining CNN-based classification with geometric heuristic verification, which meaningfully reduces false positives for visually similar gestures — a problem that pure deep learning approaches struggle with. The client-side inference model also sets a practical precedent for privacy-preserving health applications, since no biometric data (webcam frames) ever leaves the user's device.

In terms of industry applicability, the system is directly relevant to digital wellness platforms (such as apps in the yoga, meditation, and physiotherapy space), telehealth services seeking gesture-based patient monitoring, and rehabilitation tools for hand motor recovery. The underlying architecture — lightweight CNN + landmark-based heuristics + browser deployment — is generalizable to sign language recognition, hand therapy, and accessible HCI (human-computer interaction) applications. The project also holds a published patent (Application No. 202541054372 A), indicating its potential for commercialization.

Working Model vs. Simulation/Study:

This project involved the development of a fully functional working model, not a simulation or theoretical study.

The system is a live, deployable web application accessible through any standard browser with webcam access. It performs real-time hand gesture recognition and correction on actual users, processing live video frames through the complete pipeline — MediaPipe landmark detection, CNN classification, geometric heuristic verification, and temporal stabilization — all running client-side. The CNN model was trained on a real custom dataset of approximately 1,000 images collected under varied real-world conditions (different lighting, backgrounds, hand sizes, and orientations), and evaluated on actual test sessions with live users.

The MERN stack backend is also fully implemented, handling real user authentication, session tracking, and progress persistence. Performance metrics including accuracy (~92%), inference latency, and confusion matrices were derived from actual system runs, not simulated data. In summary, every component of Healing Hands was built, integrated, and validated as a working end-to-end product.

Project Outcomes and learning's:

Outcomes:

1. A fully functional browser-based mudra recognition system achieving ~92% classification accuracy across 8 mudra classes, outperforming generic CNN and Teachable Machine baselines while maintaining a significantly smaller model size.
2. A novel hybrid detection pipeline combining MobileNetV2-based CNN with Geometric Heuristic Verification, which substantially reduced false positives between visually similar mudras — a known limitation of pure ML approaches.
3. A conference paper accepted and presented, contributing to academic literature on lightweight, real-time gesture recognition systems.

4. A KSCST-funded project under the 49th Student Project Programme, recognizing its societal relevance in the domain of AI in Health.

Learning’s:

1. On the ML side: Training a CNN on a small custom dataset (~1,000 images) requires careful augmentation strategy — random rotations, brightness variation, and flipping proved critical to generalization across different users and environments. Accuracy alone is insufficient; visually similar classes need domain-specific post-processing (heuristics) to be practically reliable.
2. On system design: Client-side inference using TensorFlow.js is viable for production-grade applications, but requires deliberate model compression and architecture choice (MobileNetV2 over heavier models) to maintain acceptable frame rates on mid-range hardware.
3. On integration: Combining a stateless real-time inference engine with a stateful MERN backend required careful API design to ensure session data remained consistent without introducing latency into the detection loop.
4. On real-world deployment: Controlled lab accuracy does not directly translate to real-world performance — variations in background, lighting, and hand orientation exposed edge cases that required the temporal stabilization layer and grace interval to handle gracefully.

Scope for future work:

1. Dual-hand mudra recognition: Extend the system to support mudras requiring both hands simultaneously, enabling detection of advanced yogic practices.
2. Advanced skeletal feedback: Incorporate joint-angle analysis for more precise, posture-level corrections beyond finger-level detection.
3. Mobile application: Port the web application to a native mobile app (Android/iOS) for broader accessibility and offline usage.
4. Expanded mudra library: Increase the number of recognizable mudras beyond the current 8 classes to cover a wider range of traditional practices.
5. Full-body pose integration: Combine mudra recognition with full-body yoga pose estimation for a comprehensive AI-powered yoga assistant.
6. Personalized wellness insights: Leverage session data to generate personalized feedback, practice recommendations, and progress analytics for individual users.
7. Multi-user / classroom mode: Support simultaneous detection for multiple users in a shared session, applicable for virtual yoga classes.
8. Accessibility features: Add voice-guided instructions and screen-reader support to make the platform usable for individuals with visual impairments.