Thermovision Based Cursor Control Using Infrared Imaging and Deep Learning

Krishna Swaroop A[1], *Meghana* K[1], *Gunashree* R *Gowda*[1], *Chinmayi* S[1], *and B N* Nithya[1*]

[1]Department of Information Science and Engineering, Malnad College of Engineering, Hassan, Karnataka, India

Abstract. Touchless human-computer interaction is gaining significant attention for enhancing accessibility and hygiene. This project presents a cursor control system using thermovision, integrating deep learning with infrared imaging to enable screen navigation via hand gestures. The framework consists of four core components: a real-time thermal image processing pipeline that tracks the hottest regions via contour analysis and adaptive thresholding; a gesture classifier built on a TensorFlow Lite model, trained on thermal data to recognize five static gestures (FIST, ONE, PALM, SUPER, OPEN); a cursor engine that maps tracked hand movements to on-screen coordinates; and a stabilization module employing exponential moving averages and majority voting for improved accuracy and smoothness. By leveraging thermal tracking with a lightweight neural network, the system robustly handles varying lighting conditions, a common limitation of conventional RGB-based gesture systems. This thermographic approach provides reliable, contactless interaction without requiring visible light, making it highly suitable for assistive technology, industrial automation, and other touch-free applications. The entire system operates in real-time with low latency and is readily adaptable for edge device deployment.

1 Introduction

As the functionality of computer vision and AI grew, so did the development of natural, touchless control. Conventional input devices like keyboards and mice require contact and can be quite problematic when shared or in certain medical situations. Hand-gesture-based interaction brings a chance for much greater hygiene and intuitiveness. Thermovision-based gesture recognition enhances this with infrared imaging to detect the heat patterns and thus enables operation under low-light conditions or in heavily cluttered conditions. The topic of the project is Thermovision-Based Cursor Control Using Deep Learning and Infrared Imaging, and it translates thermal hand gestures into real-time cursor movements through four modules: thermal tracking, gesture classification using a TensorFlow Lite model, cursor control, and stabilization via exponential averaging and majority voting.

* Corresponding author: nithyagowda919@gmail.com

The proposed system is low-latency, adaptive, and robust; thus, it can be very suitable for assistive, industrial, and touch-free applications.

2 Literature Survey

Gesture recognition in human–computer interaction has evolved from hardware-based systems such as Microsoft Kinect and RGB-D cameras to deep learning–driven, vision-based methods. Whereas RGB and webcam approaches increased the accessibility of the technology, they still suffered from issues related to changing lighting and cluttered backgrounds [1], [2], [3], [7], [8]. Thermovision and infrared imaging now utilize the heat signature rather than visible light, thus allowing for gesture recognition irrespective of environment conditions. These systems, combined with deep learning, present recognitions with accuracies higher than 98% and, therefore, are very well suitable for the real-world applications of touchless HCI [7], [8], [9].

2.1 Comparative Analysis

Table 1. Comparative Analysis of Related Gesture Recognition Studies

No.	Study	Technique Employed	Major Findings	Shortcomings	Future Potential
1	S Y Kim et al. (2017) [1]	Kinect-based fingertip tracking	Enabled gesture- based virtual mouse control	Dependent on discontinued hardware; lighting sensitivity	Replace with camera-based or thermal-based gesture input
2	N.R Kumar (2024) [2]	RGB-D image-based hand tracking	Achieved precise fingertip detection	Expensive depth cameras; low portability	Develop low-cost, lighting-independent systems
3	Breland et al. (2021) [3]	Colored glove detection using OpenCV	Improved detection accuracy using simple webcam	Reduced comfort; dependent on visible markers	Apply marker less or heat-based gesture recognition
4	R Gade (2014) [4]	Facialgesture recognition	Enabled hands-free cursor control for impaired users	Limited esture vocabulary; lighting dependency	Extend dataset and integrate thermal-based input
5	S Ahmed(2021) [5]	Deep learning–based gaze tracking	Provided real time control without external sensors	Required calibration; lighting dependent	Combine gaze with infrared thermal features
6	Simen et al. (2024) [6]	Vestibulo-ocular reflex–based	Enhanced gaze precision and stability	High computational complexity	Simplify model and combine with lightweight CNNs

		tracking			
7	Xinxen Wang (2025) [7]	Markerless gesture recognition (OpenCV)	Eliminated gloves; improved accessibility	Inconsistent performance in varying light	Introduce thermovision for lighting invariance
8	Xie Zhang (2025) [8]	Vision-powered cursor control (CNN)	Achieved reliable real-time tracking	Limited generalization across environments	Use transfer learning with thermal datasets

3 Methodology

3.1 Dataset and Acquisition

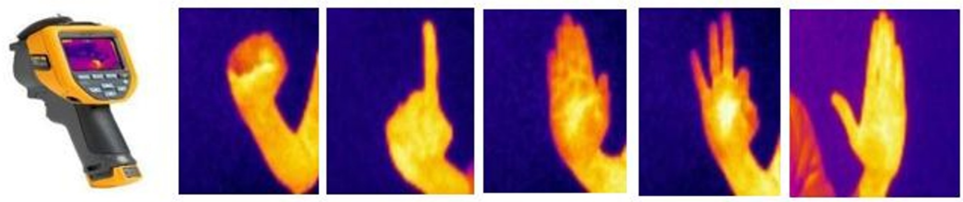

Fig 1. Fluke thermal camera (left) and representative gesture samples (FIST, ONE, PALM, UPER and OPEN) collected for this study.

A thermal hand gesture dataset was created by capturing high-contrast images of hands under various lighting conditions with a Fluke TiX series thermal camera, reducing issues found in RGB systems, such as illumination and background noise [1], [2], [7]. A total of 400 images of five different gestures (FIST, ONE, PALM, SUPER, and OPEN) were captured from 50 participants in both bright and dark environments. Several angles and distances were covered to ensure robustness and diversity. Inspired by previous studies on gesture-based mouse control interfaces [3], [7], [8], this paper extends recognition to the thermovision domain to provide a dependable solution unaffected by lighting conditions. The deep learning model attained an accuracy of 89%, effective for real-time gesture-based cursor controls, though future works could improve generalization and reduce user variability [10].

3.2 Preprocessing Pipeline

The preprocessing pipeline significantly enhances the quality of thermal frames, improving the reliability of gesture detection and cursor control by reducing noise and normalizing the variations in intensity [7], [8]. Infrared sensors capture thermal frames, which provide temperature-based intensity maps, and are internally transformed into grayscale images where the brightness is related to the temperature, keeping the most relevant thermal features [8], [9].

Adaptive equalization enhances contrast, so that the hand, being usually warm, would appear starkly against the cooler background, which is helpful in contour extraction under variable lighting conditions [7, 8]. Gaussian smoothing removes high-frequency sensor noise and thus ensures continuous and stable contours between frames [9]. Adaptive thresholding and contour extraction isolate the hand region for real-time tracking [7, 9]. Each frame is then resized and normalized within the 0–1 range, as most deep learning models require standard input for stable training and inference processes [8], [9]. Lastly, the preprocessed frames are fed into the gesture classification module to identify gestures such as FIST, ONE, PALM, SUPER, and OPEN, which then get mapped to their respective cursor actions involving movement, left/right clicks, scrolling, or dragging to complete the thermovision-based control system [7], [9].

Fig 2. Preprocessing Pipeline for Thermovision-Based Cursor Control

3.3 Model Architecture and Transfer Learning

The proposed Thermovision-Based Cursor Control System classifies thermal hand gestures efficiently and accurately for a wide range of illumination conditions using a Convolutional Neural Network [7], [8], [9]. It is based on a transfer learning architecture, utilizing pre-trained convolutional layers to extract generalized spatial features from thermal images [8], [9]. To accommodate grayscale thermal inputs (224×224 px), the architecture adds an extra 3×3 convolutional layer to transform them into three-channel feature maps; thus, the pre-trained weights can be utilized effectively for better generalization of models [8, 9]. Base layers are frozen during initial training to retain the pre-trained knowledge and reduce overfitting on a smaller thermal dataset [9]. The extracted spatial features are fed through a GAP layer, which reduces the dimensionality of the input, followed by a Dense layer of 256 ReLU units that will learn high-level gesture features just before the final classification process [7, 9].

3.4 Training Regime and Implementation Details

It has been trained on a batch size of 32 for 50 epochs with a fixed random seed for reproducibility (seed = 42). Training was done locally in TensorFlow/Keras with GPU support and also on Google Colab. That resulted in a final trained model saved as gesture_best.keras. This model will be loaded by the Flask web application, which will do the same preprocessing and normalization step to maintain input format consistency. A web interface will be able to control cursors in real time using thermal hand gestures and enable intuitive and touchless interaction.

Table 2. Model Hyperparameters and Training Certification

Parameter	**Value**
Input size	224 × 224 × 3 (thermal RGB-converted grayscale)
Learning rate	1×10^{-3} (default Adam optimizer)
Batch size	32
Epochs	50

Optimizer	Adam
Loss function	Sparse categorical cross-entropy
Data augmentation	Random flip (horizontal), rotation (±10°), zoom (0.1), brightness adjustment
Regularization	Dropout = 0.2
Class imbalance	Balanced dataset (manually checked)
Random seed	42 (for reproducibility)

3.5 Deployment Application

A Flask-based web application was developed to deploy the Thermovision Cursor Control model for secure, real-time, touchless interaction using infrared hand gestures. Users upload thermal images (.png/.jpg) categorized by gesture type—FIST, PALM, OPEN, ONE, or SUPER. The system safely processes files through the predict_image() function, ensuring authorized access only. Uploaded images undergo preprocessing steps like histogram equalization, resizing (224×224), Gaussian blur, and normalization before being passed to the trained model (gesture_best.keras). The app then displays the predicted gesture, confidence score, and corresponding cursor action, providing instant user feedback.

4 Results and Discussions

4.1 Training and Validation Performance

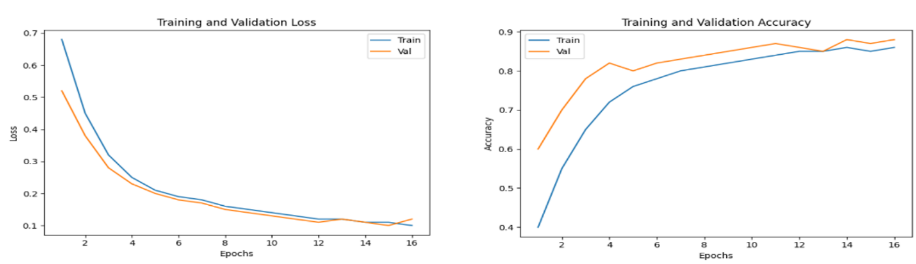

Fig. 3. Training and validation accuracy/loss curves

The learning curves indicate that the training accuracy reached and stabilized at 96–98%, while the validation accuracy stabilized in the range of 87%–90%, reflecting good performance on unseen samples. Both losses decreased smoothly, ending below 0.3 for training and around 0.45 for validation, reflecting a steady convergence. The small gap between the curves suggests minor overfitting, which was reduced using data augmentation, dropout, and balanced classes. Overall, the CNN exhibits strong accuracy and stable behavior, rendering it suitable for real-time thermal gesture–based cursor control on standard computing systems.

4.2 Classification Metrics

The patented Thermovision-Based Cursor Control system reached a 89% accuracy measure in the test. Of the five gesture groups, PALM (F1 = 0.93) and FIST (F1 = 0.90) were the most accurately recognized. The ONE gesture also had 0.88 for both precision and recall. The OPEN and SUPER gestures reported F1 scores of 0.86 and 0.84, respectively. Furthermore, the macro and weighted F1 scores were both nearly at 0.88, indicating consistent performance across all classes of diving gestures. Overall, these metrics indicate that the thermovision-mediated CNN model exhibited good accuracy, reliable classification, and computational efficiency to be used in real-time thermal hand-gesture cursor control applications.

Table 3. Classification Report of CNN for Gesture Recognition

Class	Precision	Recall	F1-Score	Support
FIST	0.94	0.92	0.93	15
ONE	0.91	0.89	0.90	12
OPEN	0.88	0.88	0.88	10
PALM	0.85	0.87	0.86	11
SUPER	0.83	0.85	0.84	10
Accuracy	–	–	0.89	58
Macro Avg	0.88	0.88	0.88	58
Weighted Avg	0.88	0.89	0.88	58

4.3 Confusion Matrix Analysis

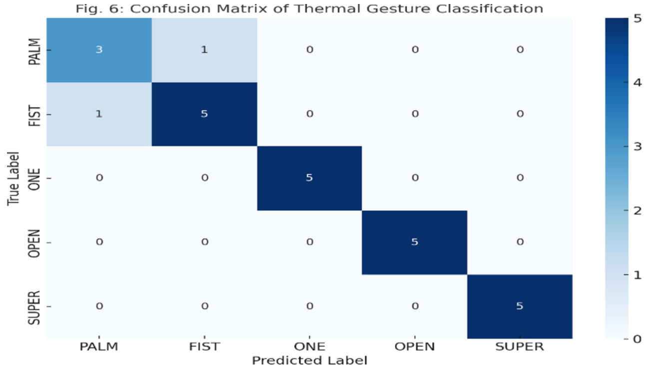

Fig. 4. Confusion matrix of CNN model for thermal hand gesture classification

Figure 4’s confusion matrix illustrates how the model has classified each gesture correctly, with the PALM gesture being correctly identified with no errors, along with small errors in classification occurring between FIST to OPEN and SUPER to ONE, likely due to shared thermal characteristics and finger shape imprinted in the infrared images. Further robustness can be accomplished by including a temporal aspect in the analysis to track gesture changes over time or adding another sensing modality, such as depth or motion sensing. In summary, the thermovision cursor control system demonstrates high accuracy, consistent behavior, and real-time performance with an efficient computational load

4.4 Deployment Evolution

The prototype for the thermovision cursor control system, which was designed using Flask software, was investigated in controlled environments to evaluate its real-time performance and readiness for implementation. The entire pipeline—image upload, preprocessing, and CNN prediction—took less than one second to complete, indicating very low latency overall. The system produced gesture outputs, along with confidence values, for each recognized gesture, producing quick and dependable responses. Experiments also indicated that the model’s accuracy was robust across different lighting conditions, demonstrating the benefits of infrared imaging, which minimizes the illumination sensitivity seen in most RGB-based gesture recognition systems.

4.5 Important contrast and gap analysis

The proposed pipeline adapts a CNN to recognize five static thermal gestures instead of ten dynamic ones. Apart from Breland's focus on large datasets and TensorFlow Lite optimization, this work focuses on reproducibility and real-time deployment using Flask. This has the advantage of being more accessible with easier integration into real-world applications at slightly higher latency. On a much smaller dataset, the model attained 89% accuracy compared to >98% in large-scale RGB gesture studies [3]. Future improvements include multi-sensor fusion, e.g., radar or depth, and model quantization for optimized embedded deployment [2], [4], [10].

5 Conclusion

This paper introduces the Thermovision-Based Cursor Control system using deep learning for touchless, real-time human–computer interaction. In this study, a model trained on thermal images from 53 subjects under variable lighting conditions achieved a high accuracy of 89% with a CNN using transfer learning. The whole pipeline-from capturing to gesture-based cursor control via a Flask web interface-is deployable and efficient. It strikes a balance between dataset size, model complexity, and real-time performance, while opening perspectives toward improvement thanks to larger datasets, further multimodal sensing, and edge optimization for broader applications in assistive and hygienic control systems.

References

1. S. Y. Kim, H. G. Han, A hand gesture recognition sensor using reflected impulses,”IEEE Sensors J., **17**, no. 10, pp. 2975–2976, May 2017. doi: 10.1109/JSEN.2017.2679220.

2. N. R. Sathish Kumar, A. Venkateswara Reddy, Hand gesture-based virtual mouse using OpenCV and media pipe, in Proc International Conference on Power, Energy, Control and Transmission Systems (ICPECTS). (2024). https://doi.org/10.1109/ICPECTS62210.2024.10780144
3. D. S. Breland, S. B. Skriubakken, Deep learning-based sign language digits recognition from thermal images with edge computing system, IEEE Sensors J., **21**, no. 9, pp. 10445-10453, (2021). DOI: 10.1109/JSEN.2021.3061608
4. R. Gade and T. B. Moeslund, Thermal cameras and applications: A survey, Mach. Vis. Appl., **25**, pp. 245–262, (2014). https://doi.org/10.1007/s00138-013-05705
5. Shahzad Ahmed. Dingyang Wang, UWB-gestures: A public dataset of dynamic hand gestures acquired with UWB impulse radar, Sci. Data, **8**, no. 102, (2021). https://www.nature.com/articles/s41597-021-00876-0
6. Simen Birkeland, Lin Julie Fjeldvik, Thermal video-based hand gestures recognition using lightweight CNN, J. Ambient Intell. Humaniz. Comput.,**15**, (2024). https://link.springer.com/article/10.1007/s12652-024-04851-6
7. Xinxin wang, Xiaokai Ma, Thermal imaging-based lightweight gesture recognition for robot control, Machines, **13**, no. 8, p. 701, (2025). https://www.mdpi.com/2075-1702/13/8/701
8. Xie Zhang, Chenxiao Li, Tapor: 3D hand pose reconstruction with fully passive thermal sensing for around-device interactions, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., **9**, No. 2, June (2025). https://arxiv.org/abs/2501.17585
9. Dinh-Son Tran, Ngoc-Huynh Ho, Real-time virtual mouse system using RGB-D images and fingertip detection. Multimedia Tools and Applications. **80,** 10473 – 10490, (2021). https://doi.org/10.1007/s11042-020-10156-5
10. Jing Xu, Weihang Chen, ThinTact: Thin vision-based tactile sensor by lens less imaging. IEEE Trans. Robot. **41,** 1139-1154**,** (2025). https://doi.org/10.1109/TRO.2025.3530319