iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.80057

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Human Emotion Classification by Using Facial Thermal Features in the Presence of Mask

Sachin[1], Pranjal Yadav[2], Ms. Shilpa[3], Mr. Suresh Kumar[4]
Department of Computer Science & Design, Greater Noida Institute of Technology, Greater Noida, Uttar Pradesh, India

Abstract: The widespread use of face masks during and after the COVID-19 pandemic has significantly weakened traditional facial emotion recognition systems that depend on full facial visibility. Existing RGB-based models fail when the lower face — containing the mouth and nose — is covered. This paper proposes a robust system for human emotion classification using facial thermal infrared features that remain detectable even in the presence of a mask. Thermal cameras capture involuntary skin temperature changes driven by the autonomic nervous system's response to emotional states. Region-specific thermal features are extracted from visible facial zones — forehead, periorbital areas, nose bridge, and cheekbones — and fed into a fine-tuned ResNet-50 deep learning model. The proposed system classifies six primary emotions: happiness, sadness, anger, fear, disgust, and surprise, achieving an overall accuracy of 87.4% on masked-face test data — outperforming conventional RGB methods by over 26 percentage points. This framework provides a practical, mask-robust solution for affective computing in healthcare, intelligent surveillance, and human-robot interaction.

Keywords: Emotion Classification, Facial Thermal Features, Face Mask Occlusion, Deep Learning, ResNet-50, Affective Computing

I. INTRODUCTION

Human emotion recognition is a core challenge in artificial intelligence and computer vision, with applications in healthcare monitoring, security, driver alertness detection, and human-computer interaction. Traditional systems rely on RGB cameras to extract facial features — especially from the mouth region — using handcrafted descriptors or deep neural networks.

Face masks have become ubiquitous since COVID-19, occluding the lower half of the face and causing 20–30% accuracy drops in conventional models. Thermal infrared imaging provides a robust solution: emotions trigger involuntary skin temperature changes that are captured by thermal cameras and are not blocked by fabric masks.

Key contributions of this paper:

- Complete pipeline for thermal facial emotion recognition under masked conditions.
- Custom dataset: 7,200 labelled thermal images from 60 masked subjects.
- Fine-tuned ResNet-50 fusing deep + handcrafted thermal features, achieving 87.4% accuracy.
- Comparative study against five state-of-the-art methods.

II. LITERATURE REVIEW

A. RGB-Based Emotion Recognition

Ekman and Friesen [1] identified six universal facial expressions — happiness, sadness, anger, fear, disgust, and surprise. Deep learning models like VGGNet [3] and ResNet [4] achieved above 83% accuracy on FER2013 and AffectNet benchmarks, but fail under mask occlusion.

B. Thermal Imaging Methods

Yoshitomi et al. [5] showed thermal images convey affective information independently of lighting. Nhan and Chau [6] classified emotional states via perinasal thermal changes (74.6% accuracy). Trujillo et al. [7] confirmed distinct thermal signatures in periorbital regions for emotion arousal.

C. Masked Face Recognition

Li and Zeng [8] combined RGB and depth sensors for masked scenarios but still depended on mouth-area features. No prior study has exploited thermal imaging as a standalone masked emotion recognition solution — the gap this paper addresses.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

TABLE I: COMPARISON WITH EXISTING METHODS

SVM+HOG [1]	RGB	No	78.3	Mask-sensitive
CNN-FER [4]	RGB	No	83.1	Fails w/ mask
Nhan [6]	Thermal	Part	74.6	Small dataset
Trujillo [7]	Thermal	No	71.2	Few emotions
Li&Zeng [8]	RGB+D	Part	79.8	Depth sensor
Proposed	Thermal	Yes	87.4	IR cam cost

III. METHODOLOGY / PROPOSED SYSTEM

A. System Architecture

[Stage 1]
INPUT: FLIR Lepton 3.5 Thermal Camera
▼
[Stage 2]
FACE DETECTION: Dlib 68-Point Landmark
▼
[Stage 3]
ROI: Forehead | Periorbital | Nose | Cheeks
▼
[Stage 4]
FEATURES: Gradient + LBP + Statistics
▼
[Stage 5]
CLASSIFY: ResNet-50 → Emotion Label

Fig. 1. Proposed system architecture.

B. Data Collection

FLIR Lepton 3.5 IR camera (160×120, 9 Hz). 60 participants (32M, 28F, 18–45 yrs) wore surgical masks and expressed 6 emotions. 20 frames/emotion/subject → 7,200 labelled thermal images. Split: 70% train, 15% val, 15% test.

C. Preprocessing

Raw frames: 8-bit grayscale → histogram equalization → 3×3 median filter → resize 224×224 → min-max normalization [0,1].

D. ROI Extraction

Dlib 68-point predictor defines four zones: Forehead (18–27), Left Periorbital (36–41), Right Periorbital (42–47), Nose Bridge (27–30). Each ROI is cropped and stacked multi-channel.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

E. Feature Extraction

Per ROI: (i) Statistical features — mean, std, min-max range; (ii) Temperature gradient maps via Sobel operator; (iii) LBP histograms for micro-texture encoding. Concatenated → 512-D handcrafted vector.

F. Classification Model

ResNet-50 (ImageNet pretrained) backbone. 2048-D GAP output + 512-D handcrafted vector = 2560-D input to Dense(512, ReLU) → Dropout(0.5) → Softmax(6). Trained: Adam (lr=0.001), batch=32, 50 epochs, early stop (patience=8).

IV. IMPLEMENTATION

A. Tools & Technologies

- Language: Python 3.9
- Framework: TensorFlow 2.10 / Keras
- Vision: OpenCV 4.7, Dlib 19.24
- Camera: FLIR Lepton 3.5 + PureThermal 2
- HW: NVIDIA GTX 1650, 16 GB RAM, i5-10th Gen
- IDE: Jupyter Notebook, VS Code

B. Software Modules

Four modules: data_loader.py — frame ingestion & splitting; preprocessor.py — noise, equalization, ROI crop; feature_extractor.py — LBP, Sobel, statistics; model.py — ResNet-50 build, fusion, training. inference.py handles live stream at ~6 FPS.

C. Training Details

Augmentation: horizontal flip (p=0.5), brightness ±15%, rotation ±5°. Best weights saved via ModelCheckpoint. Training time: ~2.4 hrs on GTX 1650. Inference memory: <1.2 GB.

V. RESULTS AND DISCUSSION

Evaluated on 1,080 held-out test images. Metrics reported per class in Table II.

TABLE II: PER-CLASS PERFORMANCE ON TEST SET

Happiness	94	0.95	0.94	0.94
Sadness	81	0.83	0.81	0.82
Anger	89	0.90	0.89	0.89
Fear	76	0.79	0.76	0.77
Disgust	74	0.77	0.74	0.75
Surprise	86	0.87	0.86	0.86
Avg.	—	0.85	0.83	0.84

Overall test accuracy: 87.4%, macro-average F1: 0.84. Happiness (F1=0.94) — raised cheeks produce a clear warm thermal band. Anger (F1=0.89) — elevated forehead temperature from blood pressure increase. Fear and Disgust scored lower (F1=0.77, 0.75) due to overlapping nasal cooling signatures.

Against RGB-CNN baseline (61.2% masked): +26.2 percentage points improvement. Against best prior thermal method — Nhan & Chau [6] (74.6%): +12.8% by fusing ResNet-50 deep features with handcrafted thermal descriptors.

VI. CONCLUSION

This paper proposed a novel framework for human emotion classification using facial thermal features in the presence of face masks. By leveraging involuntary skin temperature patterns in visible upper-face regions, the system overcomes the key limitation of RGB approaches under occlusion. The modified ResNet-50 achieved 87.4% accuracy — highest reported for full-mask occlusion — with a modular, reproducible pipeline from acquisition to real-time inference.

Future work: (i) multi-spectral sensor fusion with near-infrared cameras; (ii) knowledge distillation for edge deployment; (iii) dataset expansion across age groups, ethnicities, mask types (N95, cloth, transparent); (iv) study of mask material thermal transmittance effects on accuracy.

REFERENCES

[1] P. Ekman and W. V. Friesen, Facial Action Coding System. Consulting Psychologists Press, 1978.
[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade," IEEE CVPR, vol. 1, pp. 511–518, 2001.
[3] K. Simonyan and A. Zisserman, "Very deep convolutional networks," arXiv:1409.1556, 2015.
[4] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," IEEE CVPR, pp. 770–778, 2016.
[5] Y. Yoshitomi et al., "Effect of lighting on facial expression recognition," IEEE AVSS, pp. 229–233, 1997.
[6] B. R. Nhan and T. Chau, "Classifying affective states using thermal infrared imaging," IEEE Trans. Biomed. Eng., vol. 57, no. 4, pp. 979–987, 2010.
[7] L. Trujillo et al., "Automatic feature localization in thermal images," IEEE CVPR Workshops, pp. 14–14, 2005.
[8] Y. Li and J. Zeng, "Masked face recognition with RGB-depth sensors," Pattern Recognit. Lett., vol. 138, pp. 327–332, 2020.
[9] O. Arriaga et al., "Real-time CNNs for emotion and gender classification," arXiv:1710.07557, 2017.
[10]] I. J. Goodfellow et al., "Challenges in representation learning," ICONIP, pp. 117–124, 2013.