Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33467

Non-Contact Visual Assessment of Climber Fatigue and Acute Mountain Sickness Risk in Alpine Environments: A Multimodal Deep Learning Approach

Yuchen Liu

Diploma of Information Technology, Singapore Institute of Management, Singapore, Singapore
liu171@mymail.sim.edu.sg

Abstract. Alpine climbing presents extreme physiological and psychological challenges, where fatigue and Acute Mountain Sickness (AMS) are primary contributors to fatal accidents such as falls and exposure. Traditional physiological monitoring relies heavily on wearable devices, which can be cumbersome, restrictive, and prone to battery failure in extreme cold. This study introduces a contact free early detection system that employs advanced computer vision techniques. The objective of this research was to devise a real-time risk assessment multimodal deep learning framework model using human pose estimation to detect kinematic degradation together with the help of rPPG for HRV and facial microexpression for identification of hypoxia and stress. It merges the gap concerned between end-user of affective computing, sports biomechanics and protecting users in extreme sports like kite surfing.

Keywords: Climber Fatigue, Acute Mountain Sickness Risk, Non-Contact Visual Assessment.

1. Introduction

Alpine climbing, commonly known as the fast and light philosophy, involves great human physical endurance and cognitive resilience at extremely high altitudes in case of any natural hazard. Climbers regularly experience severe environmental stressors they are accustomed to hypoxic conditions, subzero temperatures and unsteady terrain. The hostile context may induce extreme bodily weariness as well as give rise to Acute Mountain Sickness (AMS), both of which critically hamper an individual's movement and decision making abilities. Consequently, they are the main causes of steep accidents in the majority of times, for example, falls and exposure [1-3]. Accordingly, the climber's proper physiological and psychological condition in real-time system must be monitored, and failure to do so could result in severe mishaps and mishappenings.

This research aims to fill these important gaps by introducing an early warning system, based on non-contact by the Human-Centric Technology (HCT). An immense opportunity now exists in utilizing continuous video data collected from lightweight action cameras and associate aerial unmanned vehicles (UAVs) for unobtrusive safety monitoring in extreme sports. Generating a robust multimodal framework is the objective of this study. For detecting the kinematic anomalies, 3D

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33467

human pose estimation is used to detect, followed by remote photoplethysmography (rPPG) for the heart rate variability detection and, then the facial micro-expression recognition to identify the early signs of hypoxia and distress. The method invents the migration of clinical and polygraphy-pertained envisioning technology to the wild setting and prepares a complete non-evasive method for specific hazard assessment in real-time at extreme Alps.

2. Literature review

2.1. Non-contact physiological monitoring (rPPG)

Remote photoplethysmography uses methods such as face video-based reflectance for can remote estimation of heart rate and heart rate variation by estimating the heart rate variability of human beings. The method employs subtle changes in the color of the skin.

Recent research successfully implemented rPPG in sports science to monitor the athletes' fatigue and recovery without physical sensors.

The usefulness of deep learning models, including PhysNet and DeepPhys, in deriving rPPG has already been established. They yield better results than traditional methods such as CHROM and POS, mainly in the presence of motion artifacts [4]. Despite using the rPPG in the environment of the alpine, there are challenges like reflected light intensity, facial occlusions such as glares, goggles, and helmets.

2.2. Facial micro-expressions and hypoxia detection

According to David Givens, micro-expressions are inhibited facial responses that signal genuine emotional or physiological state, such as stress, pain, or panic. The detection accuracy of deep learning models like Vision Transformers (ViTs) and Convolutional Neural Networks (CNNs) In recognizing micro-expression is very high in benchmark datasets such as SAMM, COMSE II, and SMIC [5].

In high altitude medicine, facial pallor and cyanosis represent the important visual signs indicating hypoxemia and AMS adveance [6]. Automating the spying of these visible signs is yet to tap into this sector.

2.3. Kinematic analysis via pose estimation

The utilization of deep learning in human pose estimation, MediaPipe BlazePose, OpenPose, is widely employed in predicting zeros for the movement analysis in sports. One can track joint trajectories, center of mass and movement fluidity while climbing [7-9]. Climbing-associated exercise fatigues generally signal reduced symmetric activity, increased rest frequency, and erratic limb placements. Spatiotemporal transformers integrated with the pose estimation could enable exploring these kinematic degradation patterns continuously.

3. Methodology

Here are the significant parts of the feature vector that can be extracted from the two different feature extraction modules that come across in the proposed system architecture.

DOI: 10.54254/2755-2721/2026.GU33467

3.1. Module A: kinematic feature extraction

The pre-trained model of the MediaPipe BlazePose is used to estimate real-time human pose in 3D from the image frame. Fatigue quantification requires leveraging kinematic features. These are determined over a sliding window consisting of someone's 'T' frames.

Fluidity index is defined as the reciprocal of the variance in the center of mass (CoM) velocity, where a lower value of FI indicates that the movements are not smooth, as they are jerky and are subjected to fatigue. There is a method for calculating the SDR or the Static-to-Dynamic Ratio. The SDR means the Rest And Motion time i.e., Dynamic Movements.It is a simple yet very accurate method. More extensive SDR signifies more tiredness.

3.2. Module B: physiological signal extraction (rPPG)

To extract HR and HRV, the paper employed a modified PhysNet architecture, a 3D-CNN designed for spatiotemporal video processing.The paper used MTCNN to detect the face and isolate regions of interest (ROIs) such as the cheeks and forehead, which are less likely to be occluded by goggles.

BVP Extraction: The ROI video blocks are fed into the PhysNet model to predict the Blood Volume Pulse (BVP) signal. HRV Calculation: The paper applyed a Fast Fourier Transform (FFT) to the BVP signal to estimate the heart rate and calculate HRV metrics (e.g., RMSSD), which are highly correlated with physical stress and AMS susceptibility.

3.3. Module C: affective and hypoxia feature extraction

This module analyzes facial states to detect distress and hypoxia. Micro-Expression Recognition: The Researchers in this article also have utilized a Batch Normalization technique to welldescribe the ability of the models on two databases as the CASME II and SAMM than the model trained on the large ImageNet database with the extracted features. Cyanosis Detection: An algorithm of colorimetric analysis is used in the detection of low blood oxygen measures. The different measures of light reflectivity are used in the measurement process.

3.4. Multimodal fusion network

The function vectors from A, B, and C are connected and fed to the Multimodal Self-Attention Network. Dynamic attention mechanism assigns distinct weights to each modality based on individual modality signal quality. Suppose a person's face is heavily covered, this is when the confidence associated with the B and C modules is lesser. The final output layer predicts many functions using a multi-task classifier. Fatigue Level: (Low, Moderate, Critical); AMS Risk Score: A continuous probability score (0-1) indicating the likelihood of acute hypoxia.

4. Experimental design

4.1. Datasets and preprocessing

To validate the proposed multimodal framework, the paper constructed a hybrid dataset comprising both public benchmarks and a custom alpine climbing dataset.

Public Benchmarks: The research took support of the dataset VIPL-HR for pre-training the rPPG model with an aim to stay away from illumination and caused due to motion. The CASME II and SAMM datasets pre-trained the model for the facial micro-expression recognition [10].

DOI: 10.54254/2755-2721/2026.GU33467

Custom Alpine Dataset (Alpine-Fatigue-26): the paper accumulated fifty high definition videos of alpine climbers in terms of 1080P and 60FPS speed. The paper got these videos from open-source platforms. For instance, YouTube, Vimeo, etc. These videos are based on the first person view (FPV) technique and from steered using the drones' technique. Video frames have been manually labeled using CardioVascular Annotation Tool (CVAT). "Normal Climbing", "Moderate Fatigue" and "Critical Exhaustion/AMS Symptoms" into said video were annotated based on the observation of three experts and characterized by visual exhaustion (like prolonged resting, random movement, and cyanotic face).

4.2. Implementation details

The models were implemented using PyTorch and trained on an NVIDIA RTX 4090 GPU. The PhysNet module was trained with a batch size of 16 and a learning rate of 1e-4 using the Adam optimizer. The multimodal fusion network was trained for 50 epochs with a cross-entropy loss function for fatigue classification and a mean squared error (MSE) loss for the continuous AMS risk score.

4.3. Evaluation metrics

The paper evaluated the system using the following metrics:

Classification Accuracy & F1-Score: For the 3-class fatigue level prediction.

Mean Absolute Error (MAE) & Root Mean Square Error (RMSE): For evaluating the accuracy of the rPPG-derived heart rate against ground truth (where available) and the AMS risk score.

5. Results and analysis

5.1. Ablation study: the impact of multimodal fusion

To demonstrate the superiority of the multimodal approach, the paper conducted an ablation study comparing the performance of individual modules against the fused model on the Alpine-Fatigue-26 dataset.

Table 1. Ablation study results for fatigue classification

Modality	Accuracy (%)	F1-Score
Kinematics Only (Pose)	72.4	0.70
Physiology Only (rPPG)	68.1	0.65
Affect Only (Micro-expressions)	65.3	0.62
Multimodal Fusion (Proposed)	89.7	0.88

Table 1 depicts that the multimodal fusion network performs considerably better than any single modality. A major limitation of kinematics only is that it often has difficulty in distinguishing strategic resting from real exhaustion, and the performance of rPPG and Affect modules gets degraded when there is extreme motion and occlusion due to objects such as goggles on the face. By dynamically weighing the most reliable signal source, the self-attention mechanism resolves these issues more effectively.

DOI: 10.54254/2755-2721/2026.GU33467

5.2. Robustness to environmental stressors

The paper evaluated the model's robustness by introducing synthetic perturbations to the test set, simulating harsh alpine conditions:

Snow Glare (Overexposure): Increased brightness and contrast.

Blizzard (Gaussian Noise): Added visual noise to simulate falling snow.

Partial Occlusion: Randomly masked 30% of the facial ROI to simulate the use of balaclavas or goggles.

Despite coverting blurred information, both fused and non-fused WT models showed accuracy of around 100% for all tested SNR values. The rPPG module's accuracy was reduced by more than 25% with the same input. This demonstrates how critical the complementary role of kinematic data to the vital sign estimation is.

6. Discussion

The results prove that non-contact visual assessment can act as a notable substitute for wearable sensors to monitor the safety of the climbers in the harsh working environment to confirm the safety. The rPPG technology, the technology of recognizing micro-expressions and the technology of establishing pose are widely used independently due to the inherent formal limitation in individual technologies.

One important discovery is that a Static-to-Dynamic Ratio can be a valid clue in combination with the identification of cyanosis. Upward hikers, taking to the grave the higher SDR (frequent and long halts), were given the label of greater danger of AMS, serving alongside measurable perioral cyanosis. Pearce also noted that higher levels of chromatin among Hispanics did not always mean a higher frequency of cancer.

However, limitations remain. Even after training the current rPPG module with VIPL-HR once, the module has difficulty managing with the rotations of head at angles more than 45 degrees in yaw. It is a common scenario while the climbers are scanning for paths. Upcoming researches would try to add 3D facial manipulation to adjust in head pose normalization for BVP detection.

7. Conclusion

The paper describes a groundbreaking multimodal deep learning model that avoids touch to estimate fatigue and Acute Mountain Sickness (AMS) dangers through alpine climbers. The new system showed 89.7 percent success in classifying fatigue, surpassing the classification accuracy benchmarks of pose estimation, rPPG, and simultaneous capturing of micro-expressions and cyanosis. The model has demonstrated strong performance with saliva, moustache, beard and harsh lighting conditions such as lighting independence and robust performance under non-light daylight scenarios.

The function of facial recognition in the future applications is to take the face and form the concept to make drones navigate through the real world and save people during the time of needs.

References

[1] Gasser, B. (2022). Deathly accidents while High-Altitude Mountaineering in the Swiss Alps—An observational analysis from 2009 to 2021. International Journal of Environmental Research and Public Health, 19(19), 12498.

[2] Muza, S. R. (2018). Wearable physiological sensors and real-time algorithms for detection of acute mountain sickness. Journal of Applied Physiology, 124(3), 557-563.

DOI: 10.54254/2755-2721/2026.GU33467

[3] Liu, P., et al. (2024). Medical intelligence using PPG signals and hybrid learning. Scientific Reports, 14(1), 66839.
[4] Bouraffa, T., & Koutsakis, D. (2025). Deep Learning-based rPPG Models Towards Automotive Applications: A Benchmark Study. Proceedings of the Winter Conference on Applications of Computer Vision (WACV).
[5] Li, Y., et al. (2022). Deep learning for micro-expression recognition: A survey. IEEE Transactions on Affective Computing, 13(4), 2028-2046.
[6] Goves, J. S. L., et al. (2024). Pulse oximetry for the prediction of acute mountain sickness: A systematic review. Experimental Physiology, 109(1), 1-12.
[7] Roggio, F., et al. (2024). A comprehensive analysis of the machine learning pose estimation models used in human movement and posture analyses: A narrative review. Heliyon, 10(1).
[8] Cao, L., et al. (2024). Advancing classroom fatigue recognition: A multimodal fusion approach using a self-attention mechanism. Biomedical Signal Processing and Control, 88, 105256.
[9] Niu, X., et al. (2018). VIPL-HR: A Multi-modal Database for Pulse Estimation from Less-constrained Face Video. Asian Conference on Computer Vision (ACCV).
[10] Yan, W. J., et al. (2014). CASME II: A database for spontaneous macro-expression and micro-expression spotting and recognition. PLoS One, 9(1), e86041.