

Entwicklung und Einsatz von automatisierter Drohnentechnologie für kamerabasierte  
kontaktlose Vitalparametererfassung und semi-automatisierte Triage in der  
Katastrophenmedizin

Von der Medizinischen Fakultät  
der Rheinisch-Westfälischen Technischen Hochschule Aachen  
zur Erlangung des akademischen Grades eines Doktors der Theoretischen Medizin genehmigte  
Dissertation

vorgelegt von

Lucas Mösch, M.Sc. RWTH

aus Saalfeld / Saale

Berichter: Prof. Dr. med. Dr. rer. nat. Michael Czaplik  
Univ.-Prof. Dr. med. René Tolba

Tag der mündlichen Prüfung: 05.12.2025

## Publikationsliste

### Publikationen, die die Dissertation darstellen:

**Mösch, L.**, Barz, I., Müller, A., Pereira, C. B., Moormann, D., Czaplik, M., & Follmann, A. (2023). For Heart Rate Assessments from Drone Footage in Disaster Scenarios. *Bioengineering*, 10(3), 336. <https://doi.org/10.3390/bioengineering10030336>

**Mösch, L.**, Pokee, D. Q., Barz, I., Müller, A., Follmann, A., Moormann, D., Czaplik, M. & Pereira, C. B. (2024). Automated Unmanned Aerial System for Camera-Based Semi-Automatic Triage Categorization in Mass Casualty Incidents. *Drones*, 8(10), 589. <https://doi.org/10.3390/drones8100589>

### Weitere Publikationen:

Roecher, E., **Mösch, L.**, Zweerings, J., Thiele, F. O., Caspers, S., Gaebler, A. J., Eisner, P., Sarkheil, P. & Mathiak, K. (2024). Motion Artifact Detection for T1-Weighted Brain MR Images Using Convolutional Neural Networks. *International Journal of Neural Systems*, 34(10), 2450052-2450052. <https://doi.org/10.1142/s0129065724500527>

**Mösch, L.**, Kunczik, J., Breuer, L., Merhof, D., Gass, P., Potschka, H., Zechner, D., Vollmer, B., René, T., Häger, C., Bleich, A., Czaplik, M. & Pereira, C. B. (2023). Towards substitution of invasive telemetry: An integrated home cage concept for unobtrusive monitoring of objective physiological parameters in rodents. *Plos one*, 18(9), e0286230. <https://doi.org/10.1371/journal.pone.0286230>

Breuer, L., **Mösch, L.**, Kunczik, J., Buchecker, V., Potschka, H., Czaplik, M., & Pereira, C. B. (2023). Camera-Based Respiration Monitoring of Unconstrained Rodents. *Animals*, 13(12), 1901. <https://doi.org/10.3390/ani13121901>

Kunczik, J., Hubbermann, K., **Mösch, L.**, Follmann, A., Czaplik, M., & Barbosa Pereira, C. (2022). Breathing pattern monitoring by using remote sensors. *Sensors*, 22(22), 8854. <https://doi.org/10.3390/s22228854>

Queirós Pokee, D., Barbosa Pereira, C., **Mösch, L.**, Follmann, A., & Czaplik, M. (2021). Consciousness detection on injured simulated patients using Manual and Automatic classification via visible and infrared imaging. *Sensors*, 21(24), 8455. <https://doi.org/10.3390/s21248455>

Wienhold, J., **Mösch, L.**, Rossaint, R., Kemper, I., Derwall, M., Czaplik, M., & Follmann, A. (2021). Teleconsultation for preoperative evaluation during the coronavirus disease 2019 pandemic: a technical and medical feasibility study. *European Journal of Anaesthesiology| EJA*, 38(12), 1284-1292. <https://doi.org/10.1097/EJA.0000000000001616>

Meyer-Baese, L., Roecher, E., **Moesch, L.**, Bzdok, D., & Mathiak, K. (2020, May). CNN to detect differences in cerebral cortical anatomy of left-and right-handers. In *Real-Time Image Processing and Deep Learning 2020* (Vol. 11401, pp. 30-35). SPIE. <https://doi.org/10.1117/12.2558938>

## Article

# For Heart Rate Assessments from Drone Footage in Disaster Scenarios

Lucas Mösch <sup>1,\*</sup>, Isabelle Barz <sup>2</sup>, Anna Müller <sup>1</sup>, Carina B. Pereira <sup>1</sup>, Dieter Moormann <sup>2</sup>, Michael Czaplik <sup>1</sup> and Andreas Follmann <sup>1</sup>

<sup>1</sup> Department of Anesthesiology, Faculty of Medicine, RWTH Aachen University, 52074 Aachen, Germany

<sup>2</sup> Institute of Flight System Dynamics, RWTH Aachen University, 52062 Aachen, Germany

\* Correspondence: lmoesch@ukaachen.de

**Abstract:** The ability to use drones to obtain important vital signs could be very valuable for emergency personnel during mass-casualty incidents. The rapid and robust remote assessment of heart rates could serve as a life-saving decision aid for first-responders. With the flight sensor data of a specialized drone, a pipeline was developed to achieve a robust, non-contact assessment of heart rates through remote photoplethysmography (rPPG). This robust assessment was achieved through adaptive face-aware exposure and comprehensive de-noising of a large number of predicted noise sources. In addition, we performed a proof-of-concept study that involved 18 stationary subjects with clean skin and 36 recordings of their vital signs, using the developed pipeline in outdoor conditions. In this study, we could achieve a single-value heart-rate assessment with an overall root-mean-squared error of 14.3 beats-per-minute, demonstrating the basic feasibility of our approach. However, further research is needed to verify the applicability of our approach in actual disaster situations, where remote photoplethysmography readings could be impacted by other factors, such as blood, dirt, and body positioning.

**Keywords:** contactless; mass-casualty incident; drones; RGB video; signal; extraction; triage



**Citation:** Mösch, L.; Barz, I.; Müller, A.; Pereira, C.B.; Moormann, D.; Czaplik, M.; Follmann, A. For Heart Rate Assessments from Drone Footage in Disaster Scenarios. *Bioengineering* **2023**, *10*, 336. <https://doi.org/10.3390/bioengineering10030336>

Academic Editors: Mark van Gastel, Osama Mazhar, Wim Verkruijsse, Liang Luo and Rossana Madrid

Received: 24 January 2023

Revised: 26 February 2023

Accepted: 2 March 2023

Published: 7 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The possibilities for using unmanned aerial systems (UAS) or drones in health emergencies have increased [1]. UASs can cover short-to-medium distances in a relatively short time and access locations that are difficult for rescuers to reach [2,3] while avoiding environmental hazards [4,5]. During a mass-casualty incident (MCI), the situation on-site, the number and location of injured people, and the possible evacuation routes have been evaluated successfully using drone-camera footage [6]. The systems used ranged from conventional cameras, with live feeds to aid in decision-making [7], to thermal cameras, used to locate missing people [8], and 3D-camera setups for reconstructing a disaster scene [9].

The current procedure during MCIs is for responders to physically evaluate injured persons to assess their condition. This assessment is part of the triage process, where the severity of their injuries categorizes the injured. However, depending on the accessibility of the scene, it can take from a few minutes up to several hours for emergency personnel to locate and categorize injured people. In order to assess their condition, basic vital signs, such as heart rate (HR) and respiratory rates (RR), are measured. The gold standard for assessing HR is palpation, which requires direct contact with the victim.

The contactless assessment of vital parameters via drones could save crucial time, as obstacles do not impair drones. Furthermore, UAVs have the potential for faster and more efficient assessments of individual health states in remote and hard-to-reach locations. Our presented research aimed to investigate the technical possibilities, the capabilities, and the limitations of drone-based heart-rate assessments in order to provide a basis for future drone-assisted triage.

Several research publications have shown that camera-based HR measurements are feasible [10,11]. The developed algorithms were designed for different scenarios, such as monitoring pre-term babies in neonatal intensive care units [12], monitoring patients in intermediate care units [13], detecting infections (COVID-19) [14], and monitoring drivers [15].

Image photoplethysmography (iPPG), also called remote photoplethysmography (rPPG) or distance photoplethysmography, is a widely used method for the camera-based assessment of HRs [16,17]. Similar to contact-PPG, rPPG measures variations in the intensity of the light reflected by the skin [18]. The environmental light and the video camera substitute the light source and photo-detector, respectively. The rPPG signal is then computed from consecutive video frames, primarily using facial regions-of-interest (ROIs), such as the forehead and cheeks [19,20]. The main challenge in estimating rPPG signals with camera systems is the extremely low signal strength of the pulse-related intensity changes, as compared to those caused by environmental lighting or system motion. In order to obtain a reliable HR, several different methods were combined via multi-stage algorithmic pipelines. The current research focused on three main methods:

- (i) Selection and temporal tracking of relevant facial regions;
- (ii) Modeling of the recorded signals for PPG signal extraction;
- (iii) De-noising of rPPG signals.

Regarding the selection of facial regions and temporal tracking (i), Kumar et al. proposed a method using a Haar-cascade-based facial detection [21] and a Kanade–Lucas–Tomasi (KLT) feature tracker [22,23] to identify and track candidates with suitable rPPG regions. Those regions were then further weighted to reduce the impact of regions that could contain noisy signals. For modeling the pulse-related color changes (ii), there are two standard models, namely CHROM (Chrominance), proposed by De Haan et al. [24], and POS (Plane Orthogonal to Skin), proposed by Wang et al. [25]. Both models use a pre-defined skin-color vector with an adaptive whitening to separate the recorded RGB signals into an rPPG signal component and an orthogonal noise component. To further improve their approach, Wang et al. added a de-noising step (iii) to their rPPG model, and called their result discriminative signatures [26]. For de-noising, the authors used generalized linear modeling (GLM) to remove the first-order components of known noise signals from the recorded PPG signals. As the only known noise source for these signals, the motion recorded by the facial landmark tracking was used.

Although current approaches have performed considerably well, they also required a highly controlled environment with minimal interference to guarantee reliability. They could only extract reliable vital parameters when a person moved as little as possible. Furthermore, they were prone to changes in lighting conditions and variations in capturing angle. The utilization of these methods in outdoor scenarios poses the challenge of mitigating the effects of uncontrollable ambient lighting on the obtained data. Additionally, when utilizing drone-mounted systems, an additional difficulty is introduced in order to compensate for the system's increased motion and the increased distance between the system and the subjects being monitored. These challenges had to be addressed to ensure the obtained data's accuracy and reliability.

This paper proposed a novel algorithmic pipeline specifically developed and optimized for estimating HRs from videos captured by a UAS. The primary objective of this research was to evaluate the potential of UAS-based rPPG as a first step in the assessment of the physiological status of victims in MCI by leveraging the capability of UAS to reach the disaster location in a shorter time than first-responders. As a secondary objective, we investigated the selection of appropriate de-noising sources to compensate for the system movement and to study the effects of environmental lighting on HR assessments from rPPG signals. Additionally, we provided a full implementation of this approach, along with a proof-of-concept study that could improve patient outcomes and incident scene management.

## 2. Materials and Methods

### 2.1. Imaging Pipeline Overview

The following sections present the estimation of vital parameters via drone videos. The concept consisted of five steps:

- (i) **Data acquisition and image pre-processing:** Adjustments made to the system's parameters before and during the measurements;
- (ii) **ROI selection and tracking:** Selection of beneficial facial regions to track throughout the recording to extract rPPG signals;
- (iii) **rPPG signal extraction:** Combining the measured raw data from which the rPPG signals would be extracted;
- (iv) **De-noising and post-processing:** Refinement of the extracted rPPG signals to obtain a predominantly sparse signal for assessing HR;
- (v) **Heart-rate estimation:** Final step of calculating the HR from the extracted rPPG signals.

### 2.2. Data Acquisition and Image Pre-Processing

Data acquisition and pre-processing were the crucial parts of the pipeline, as any deficiencies would directly impact all subsequent steps. Our goal was to obtain the most stable possible recording of an individual's head. The recording was conducted on the raw data format because video encoding with common H.264 (advanced video coding) or H.265 (high-efficiency video coding) encoders could impair the subsequent HR assessments [27].

#### 2.2.1. Face-Aware Adaptive Exposure Time Adjustment

The intensity of the rPPG signals was proportional to the skin-reflected light [25]. The greater this intensity, the higher their ratio, as compared to the ambient and sensor noise (signal-to-noise ratio, SNR). The amount of light captured could be controlled through the camera's exposure time. Integrated exposure time algorithms usually attempt to generate evenly illuminated recordings. However, we experienced that integrated algorithms exposed facial regions either too much or too little. Therefore, we developed a software-driven exposure algorithm that maximized the exposure of facial regions by applying facial detection before the recording.

The images provided by the camera were denoted as  $I = [i_{t=0}, \dots, i_N]$ , and the corresponding exposure time was  $L = [l_{t=0}, \dots, l_N]$ . Empirical evidence showed that to improve the SNR of the pulse signal, the 98th percentile of the pixel values within the face region  $v$  had to be between 75% and 90% of the gray-scale range, i.e., between the pixel values 190 and 230, for an 8-bit image. The goal was to obtain as much exposure as possible but, at the same time, prevent signal clipping. For obtaining the region containing the face-related pixels, pre-existing facial recognition approaches based on Haar cascades [21], HOG + linear SVM [28], fast R-CNN [29], and YOLO [30] were tested. The facial recognition algorithm had to operate on the hardware integrated into the drone, which had very limited computing power. Therefore, simple Haar cascades were used since they enabled sufficient facial recognition at a higher computational speed. Since the relationship between exposure time  $e_n$  and facial pixel would not be known beforehand, we used the iterative adjustment algorithm (Algorithm 1) after using the cameras initially provided  $l_0$ .

---

**Algorithm 1** Automatic exposure time adjustment with respect to a person's face

---

```

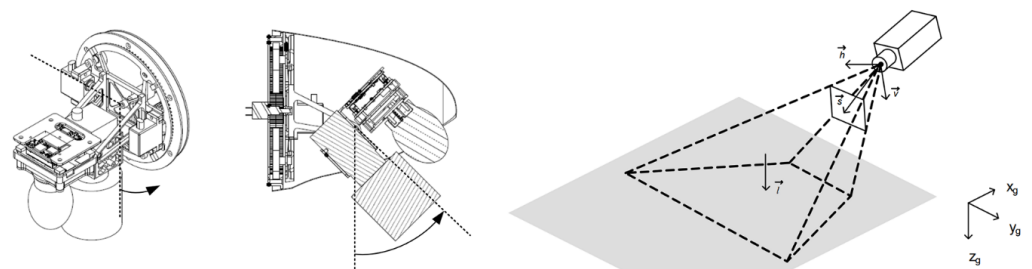
1: let  $i_0$  the initial image frame as gray image
2: let  $bb_0$  the initial facial bounding box within  $i_0$ 
3: let  $l_0$  the initial exposure time
4: for  $n = 0$  to 10 do
5:    $v = \text{percentile}(\text{pixel}(bb_n), 98)$ 
6:   if  $190 < v \text{ s.} < 230$  then
7:     return  $l_n$ 
8:   else
9:      $l_{n+1} \leftarrow \frac{190+230}{2v} \cdot l_n$ 
10:     $l_{n+1} \leftarrow \text{image}(l_{n+1})$ 
11:     $bb_{n+1} \leftarrow \text{haarcascade}(l_{n+1})$ 
12:   end if
13: end for
14: return  $l_n$ 

```

---

### 2.2.2. Active Image Stabilization

During operation, drones have to withstand a variety of weather conditions. Depending on the wind gusts, the position relative to the target could dynamically change. The deviations could vary by several centimeters, thus creating additional movement, vibration, and shake during the recordings. Electromechanical gimbals compensated by maintaining a constant visual axis relative to the target. For this purpose, the gimbals used a motor-driven cardanic suspension that immediately compensated for the smallest movements with the assistance of acceleration sensors and an active control system. With a three-axis gimbal, only perspective distortions occurred during the recordings of the mounted cameras. Since we were restricted to using a two-axis gimbal that could only compensate the pitch and roll axes, the captured images were additionally rotated by a certain degree, which required additional compensation. For this, we used the view-axis related to the left-facing vector  $\vec{h}$  and the down-facing vector  $\vec{v}$  in geodetic coordinates provided by the gimbal, as shown in Figure 1.



**Figure 1.** (Left) Technical drawing of the available gimbals in perspective and side views. Shown is the suspension of the sensors and the movable axes. (Right) The view frustum generated by the camera gimbals and the view vectors.

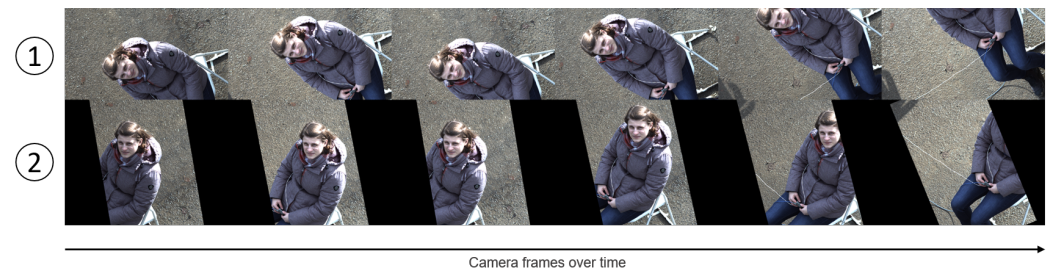
As a first step, the synchronized pairs of the gimbals and image data were created, as the acquisition of both was usually not synchronous and performed at different frequencies. For this, we used the provided acquisition timestamps  $t_n^i$  of the images and  $t_m^g$  of the gimbal vectors. For each image–timestamp pair  $(i_n, t_n^i)$ , the corresponding gimbal-vectors  $(\vec{h}_m, \vec{v}_m, t_m^g)$  were determined by minimizing the temporal distance:

$$|t_n^i - t_m^g| \rightarrow \min. \quad (1)$$

With the corresponding gimbal vectors, the rotation of the optical plane  $y_n$  of each image could be calculated using the z-component of  $\vec{h}_m, \vec{v}_m$  by:

$$y_n = \text{atan2}(h_m^z, v_m^z). \quad (2)$$

Afterward, each image  $i_n$  was digitally rotated by  $-y_n$  using bilinear interpolation to remove the image rotation caused by the gimbal. All images were centered within a zero-padded quadratic frame of a squared-image width to maintain a rectangular image shape. Therefore, a new set of quadratic, rotated images  $I_{rot} = [i_{rot,t=0}, \dots, i_{rot,N}]$  of equal size were created, as demonstrated in Figure 2.

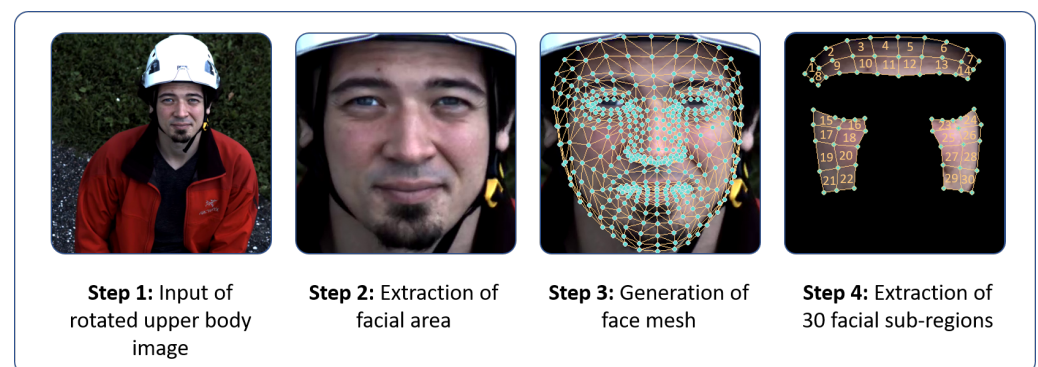


**Figure 2.** Time series of recorded camera frames before (1) and after (2) image stabilization.

### 2.3. ROI Selection and Tracking

The next step of our approach was to process the rotation-free images further by identifying and tracking the facial ROIs. Figure 3 displays the four major steps used for this purpose.

First, a facial bounding box within each image of  $I_{rot}$  was calculated to extract the facial area (Figure 3, Step 2). The entirety of the facial areas resulted in a time series of the facial size and position  $B = [bb_{t=0}, \dots, bb_N]$ . Here, the same Haar-cascade-based classifier used for the adaptive exposure time adjustment was used for the calculation. However, due to the movements of the UAS, only some images could contain a face entirely; thus, face detection inevitably failed when this was not the case. As a result, the time series could be discontinuous, creating several time-frames of consecutive facial areas  $B_{t=[a,b]}, B_{t=[a',b']}, \dots$  for the same recording. Therefore, all time-frames of facial areas that were at least five seconds long were used for further processing. In cases where no such time-frame existed, the assessment of the HR was declared as failed. Within the facial area of each time-frame, the facial ROIs required for the rPPG extraction were determined using a 3D-mesh representation of the face (Figure 3, Step 3). For this purpose, the face mesh function of Google's Media Pipe framework was used [31,32]. The provided function combined a neural network for facial feature extraction with a feature-based 3D transformation to generate 468 3D landmarks from 2D facial images in order to create a face mesh. In the end, 30 sub-regions covering the forehead and cheek regions were selected within this face mesh. They represented the ROIs used for extracting rPPG signals. Finally, the raw rPPG signal was determined for each ROI by averaging the pixel color values within the region throughout the video frames of the time series.



**Figure 3.** Individual steps for facial ROI selection and tracking.

#### 2.4. De-Noising and Signal Post-Processing

The raw signals extracted from the different ROIs were usually too noisy to be used as rPPG signals. In general, the noise was related to movements caused by wind and wind changes and the inherent vibrations of the drone, for which the gimbal could not fully compensate. It was also related to the perspective distortions and changing lighting conditions. Therefore, a generalized linear-model-based de-noising was developed to reduce the noise caused by the UAS and the environmental changes in lighting. This de-noising consisted of modeling the raw signal of each color channel and each ROI  $y_R, y_G, y_B$  as the composition of a linear mixture of known noise source signals  $X$  through the mixing matrix  $m$ . The blood flow correlated to the rPPG signal  $p$  and the remaining unknown noise component  $e$ , as expressed by:

$$\begin{aligned} y_1 &= m_1x_{1,1} + m_2x_{2,1} + \dots + p_1 + e_1 \\ y_2 &= m_1x_{1,2} + m_2x_{2,2} + \dots + p_2 + e_2 \\ y_3 &= m_1x_{1,3} + m_2x_{2,3} + \dots + p_3 + e_3 \\ &\vdots \\ y_n &= m_1x_{1,n} + m_2x_{2,n} + \dots + p_n + e_n, \end{aligned} \quad (3)$$

where  $y_1 \dots y_n$  denotes the temporal components of an ROI's color component and  $m_1x_1, m_2x_2, \dots$  denotes the weighted components known source of noise. Assuming an equally known source of noise for each color channel, this resulted in:

$$\begin{aligned} y_R &= X \cdot m_R + p_R + e_R \\ y_G &= X \cdot m_G + p_G + e_G \\ y_B &= X \cdot m_B + p_B + e_B. \end{aligned} \quad (4)$$

To determine the noise contribution components of  $m$ , the former equation could be rewritten as:

$$Y = X \cdot M + P + E. \quad (5)$$

To solve for  $M$ , a standard least-squares approach for over-determined systems was chosen. With this contribution matrix, the raw signal  $Y$  could be de-noised by subtracting the noise contributions:

$$\hat{Y} = Y - X \cdot M = P + E, \quad (6)$$

where  $\hat{Y}$  represents the de-noised signals. Since the de-noising performance of this approach depended solely on the selection of appropriate known noise sources, their selection was crucial and had to satisfy two criteria. On the one hand, the noise sources had to correlate with the actual noise influence within the raw signal. On the other hand, they could not correlate with the blood-flow-related changes, as this removed them during the de-noising process. In addition, too many noise sources could ruin the de-noising process, as the accumulated noise could also lead to the removal of blood-flow-related signal components.

For our approach, 15 sources of noise were used, as listed in Table 1. To compensate for noise components correlated with the unwanted movement of the UAS, data provided by the flight systems sensors were used, namely from the GPS and the accelerometer. They were provided and pre-processed identically to the gimbal data by matching them to the corresponding timestamps. Therefore, matching time series were created from 6 noise source signals that included (1) **height above ground**, (2) **longitude**, (3) **latitude**, and (4) the (u,v,w) values of the **velocity vector** in world coordinates. In addition, the time series of the (x,y,z) values of the two gimbal vectors (5)  $\vec{h}$  and (6)  $\vec{v}$  were included as noise source signals, as well.

For compensation of the movement correlated with the tracking of facial regions, the (x,y) values of the (7) **landmarks** forming the facial ROI and their (8) **center-of-mass** were utilized. The number of landmarks used varied between four and six, depending on the ROI. Moreover, the (9) **number of pixels** included in the facial ROI was used as

well. Finally, the mean color values of the facial regions not containing blood-related color changes were utilized to account for other noise influences. These were affected by other noise influences, such as changes in lighting, in the same way as the facial ROIs but did not contain blood-related color changes. Therefore, they could be used as direct noise sources that accounted for these influences. For this, the sub-regions covering the (10) **nostrils** and the (11) **eyes** were processed in the same way as the facial ROIs by using the time series of their mean color values. Thus, four additional noise source signals were created for de-noising.

The linear modeling of the de-noising problem removed only linearly correlated components. To consider higher-order influences, we also used the first-order derivatives of the (12) **landmarks**, (13) **center-of-mass**, and (14, 15) **gimbal vectors**. As a result, depending on the number of landmarks creating the sub-regions, 42 to 50 individual signals were created and used for de-noising.

**Table 1.** List of known 15 noise sources and their individual components that were used in the de-noising process.

|                                      |                                     |                                |
|--------------------------------------|-------------------------------------|--------------------------------|
| landmarks (4–6) (x,y)                | center-of-mass (x,y)                | number of pixels               |
| $\frac{d}{dt}$ landmarks (4–6) (x,y) | $\frac{d}{dt}$ center-of-mass (x,y) | height above ground            |
| longitude                            | latitude                            | velocity (u,v,w)               |
| nostrils (left, right)               | eyes (left, right)                  | $\vec{h}$ (x,y,z)              |
| $\vec{v}$ (x,y,z)                    | $\frac{d}{dt} \vec{h}$ (x,y,z)      | $\frac{d}{dt} \vec{v}$ (x,y,z) |

With the matrix  $M$  consisting of these known noise signals, every color channel of the 30 sub-ROIs was de-noised independently, creating a new set of de-noised time-frame signals. The last part of this step consisted of generating an rPPG signal for each region and time-frame, i.e., continuous video sequences with a length of at least 5 s. For this purpose, the individual color signals had to be combined to form an rPPG signal. In the literature, several methods existed. The more sophisticated approaches aimed to obtain the blood-flow-related components through color channel mixing. This work applied two approaches: POS [25] and CHROM [24]. Both algorithms attempt to determine the color vector most closely corresponding to the blood-related color changes by estimating the face color in order to project the color signals onto it, using the variance of the power spectrum of the changes. In our approach, both rPPG signals created by POS and CHROM were calculated and used for robust HR assessments.

## 2.5. Heart-Rate Estimation

Most of the current rPPG approaches focus on the continuous monitoring of patients, and consequently, they focus on detecting variations in HR. However, in disaster medicine, the temporal variation of this parameter can be less relevant than the immediate assessment of an average value for patient triage. Therefore, our approach focused on obtaining a single value for HR.

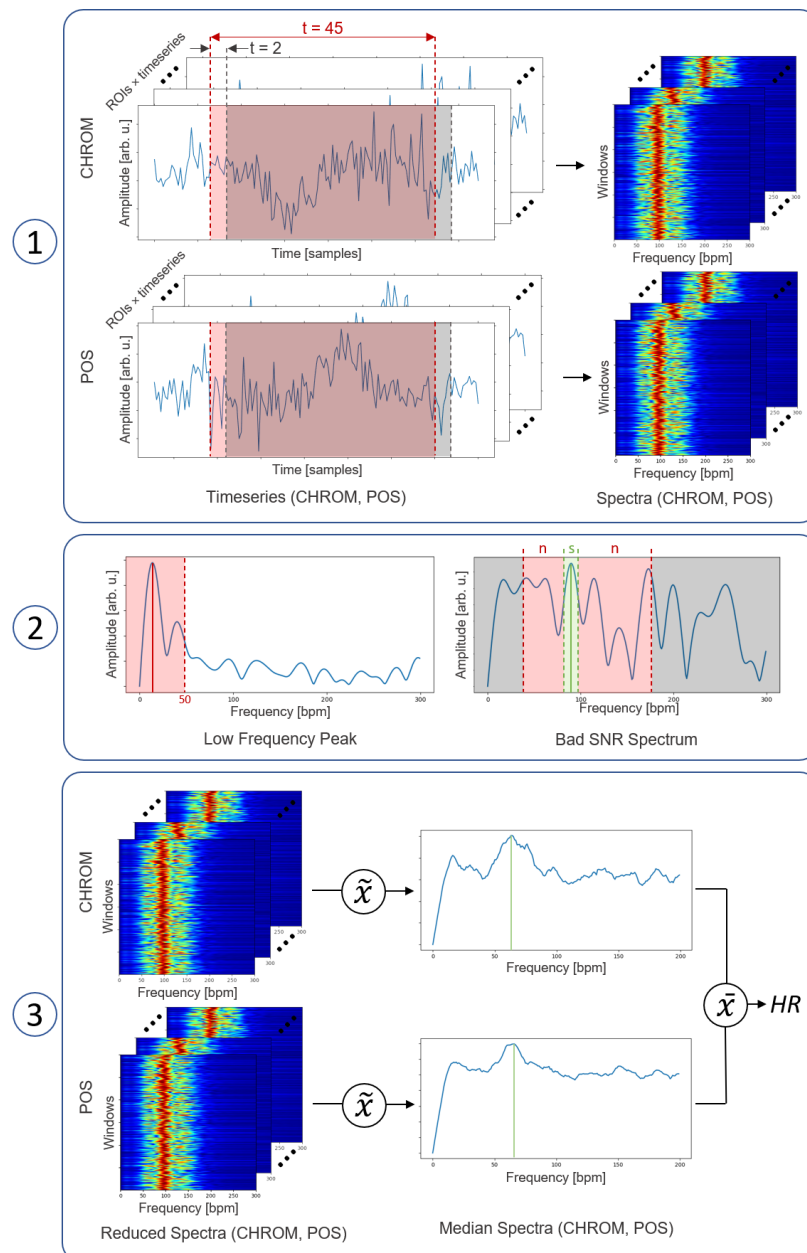
For an HR assessment, it was assumed that an HR fluctuated only slightly over a short period. Hence, equivalent blood-related changes could be measured at any time and in any facial ROI with both color-mixing methods. In Figure 4, we detailed how this was utilized.

First, each de-noised ROI signal was sliced into overlapping sub-signal windows of 45 samples (i.e., 5 s) with a stride of 2 samples, and the frequency spectrum  $P_n$  of each was calculated using the Fourier transform (Figure 4, (1)). Second, improper sub-signal spectra were removed from this set of spectra. Each spectra had to satisfy two conditions to be considered for HR estimation (Figure 4, (2)). As a first condition, any frequency peak within the target range of 50 bpm (beats-per-minute), up to 180 bpm, had to have an SNR

of at least 2.5%. For this, a prominence-based peak search was conducted within the target frequency range, followed by the SNR calculation for each peak  $p$ :

$$SNR = \frac{\sum_{p-\alpha}^{p+\alpha} P(p)}{\sum P - \sum_{p-\alpha}^{p+\alpha} P(p)}. \quad (7)$$

For the peak width  $\alpha$ , a value of 3 bpm was chosen. By adding this criterion, we intended to exclude spectra that were too noisy from the HR assessment. As a second condition, there had to be no global maximum regarding the target range within the low-frequency range between 0 bpm and 50 bpm. This condition prevented the higher harmonics produced by prominent low-frequency noise components from concealing the HR-related frequency within the target range.



**Figure 4.** Overview of the HR estimation process—Windowing, time-series creation, and spectra calculation (1). Rejection of outlier spectra with low-frequency components or insufficient SNR (2). Final HR assessment by median spectra creation (3).

Third, all remaining spectra were combined for CHROM and POS, via a median, resulting in two total spectra each. A resulting HR value was then determined for both total spectra by determining the frequencies corresponding to the respective maximums in the range of the target range. The final HR value was then calculated by the mean value of the two HR values determined via POS and CHROM. The mean value was chosen because if the HR values of the two methods differed, it could not be determined which was correct. By averaging, the values correctly recorded by one method could worsen. However, the substantial deviations from the actual value were reduced, making the robust detection of a single value more valuable.

## 2.6. Experimental Evaluation

### 2.6.1. Flight System and Camera Setup

The drone used for the experimental study was a tilt-wing UAS developed by the Institute of Flight System Dynamics (FSD), RWTH Aachen University, and the flyXdrive GmbH (FXD), in Aachen, Germany. It was based on the “Neo” tilt-wing UAS manufactured by FXD and was provided within the FALKE research project (funded by the Federal Ministry of Education and Research).

Tilt-wing aircraft are characterized by their ability to perform efficiently and fast forward-flight for bridging longer distances as well as hover-flight, allowing vertical takeoff and landing almost anywhere. Depending on the flight mode, an operation time of up to one hour can be achieved, whereby in pure hovering flight, the system would be limited to a shorter operation time. During the fixed-wing flight, a top speed of up to 130 km/h could be achieved. During the research project, the drone, displayed in Figure 5, was further adapted by the consortium to include sensors, and a gimbal was required towards the front of the aircraft for the remote assessment of vital signs. The 2-axis gimbal was integrated into the system to compensate for system motion during video recordings. This could be controlled on the roll-axis, from  $-90^\circ$  to  $+90^\circ$ , and the pitch-axis, from  $-15^\circ$  to  $+60^\circ$ , with an actuation speed of  $400^\circ/\text{s}$ .

The UAS was equipped with an Allied Vision Mako G-234C RGB (Allied Vision Technologies GmbH, Stadtroda, Germany) camera utilizing a Sony IMX249 CMOS sensor (Sony Group Corporation, Tokyo, Japan). This was chosen because it had a volume of 28 ccm, weighed 80 g, could capture raw 10-bit Bayer frames at up to 41.2 frames-per-second (fps), and had a special resolution of up to 1936-by-1216 pixels. In addition, it had a typical C-mount that could be used to mount any optics that could enable it to function at various distances.



**Figure 5.** 3D rendering of the “Neo” tilt-wing UAS by flyXdrive used in the study.

### 2.6.2. Data Acquisition

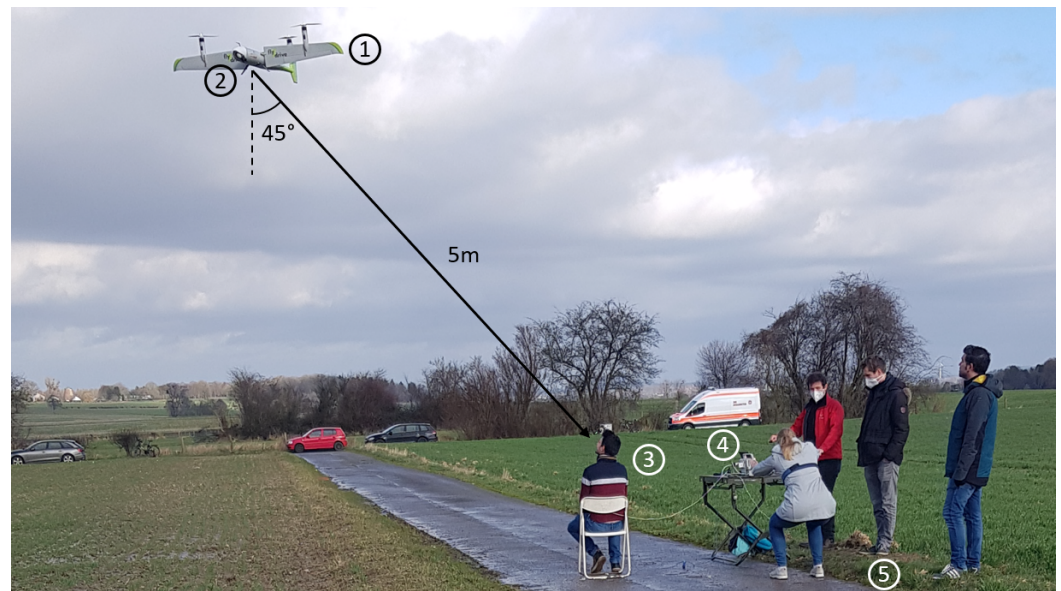
For the recordings, the camera was equipped with a Kowa 50 mm 1” 5MP C-mount lens (Kowa Optimed Deutschland GmbH, Düsseldorf, Germany). Frame capturing was conducted in an 8-bit Bayer format of 1920-by-1080 pixels at a frame rate of 15 fps. For managing and storing the recordings, a NanoPC-T4 (open-source) operating a custom recording software was integrated into the UAS.

### 2.6.3. Experimental Setup and Study Protocol

To experimentally evaluate our system for HR assessments, a group of 18 subjects was recorded with the UAS during flight under different conditions. The experiments were conducted with the assistance of healthy volunteers. The ethics committee of the medical faculty of the RWTH Aachen University in Germany approved the conduct of the study. Informed consent was obtained from all subjects involved in the study. The group of volunteers consisted of 6 females and 12 males between the ages of 20 and 36, with mainly white skin color.

As the UAS required a special ascent permit, the study had to be conducted in an open field that was available for FSD test flights. Since there was no other infrastructure, especially for charging the UAS's batteries and connecting electronic hardware, the study required a significant amount of organizational and personnel effort. Therefore, all recordings of the study were performed on a single day in winter.

The general setup of the recording is shown in Figure 6. The aim was to record the volunteers at a viewing distance of 5 m and at a horizontal angle of  $45^\circ$ . Since no live view was available during the recordings, a fixed position was measured for the volunteers in advance of the recordings, which the UAS permanently targeted during the flight. For each recording, the volunteers were asked to sit in the same exact position (Figure 6, (3)). Two consecutive recordings were captured for each subject. These recordings were referred to as Phases I and II. Recording time was limited to 30 s to better accommodate the drone's limited flight time and the triage aspects, as rapid assessments were a priority. To simulate different HRs, the 18 volunteers were divided in three groups of 6 each. In the first group, resting HR conditions were considered. In the second and third groups, the aim was to simulate "abnormal" physiological conditions. To achieve higher than usual HRs during Phase I and eventually Phase II, group two was instructed to exercise heavily immediately before recording in order to simulate tachycardia. Group three was instructed to breathe as slowly as possible during Phase I to lower the HR gradually. The recordings were conducted during the daytime.



**Figure 6.** Experimental setup: UAS (1) equipped with gimbal and camera (2) was hovering within a 5 m viewing distance and target engagement angle of  $45^\circ$  of a volunteer subject at a fixed position (3). The reference pleth signal was recorded by a patient monitor and stored on a computer (4). The rest of the respective subject group was preparing for their recordings (5).

For validation purposes, the vital parameters of the volunteers were assessed with a patient monitor, Philips MP2 (Koninklijke Philips N.V., Amsterdam, The Netherlands). Only the pulse oximetry was recorded at a sampling rate of 100 HZ.

### 3. Results

#### 3.1. Recording Conditions

The recordings were conducted during daylight between 11 a.m. and 4 p.m. During the recordings, the external weather conditions were harsh. The average wind force on the day of the recording was 6–7, corresponding to a wind speed of about 14 m/s. In addition, the weather was highly dynamic between sun, clouds, and light rain, which resulted in constantly changing lighting conditions. The outside temperature was about 7 °C.

#### 3.2. Target Acquisition and Adaptive Exposure Time Adjustment

The ability of the UAS to maintain the target acquisition area and the subsequent adjustments of the exposure time was evaluated based on its ability to provide appropriate time series of more than 15 s (i.e., 225 samples) as input for the presented approach.

The gimbal could not always compensate for the UAS displacement caused by the gusty winds on the recording day, as the wind pushed the UAS up to 80 cm away from its designated position. Due to these uncompensated shifts, the most crucial area-of-interest, the face, was entirely or partially lost during the recording. The complete or partial loss of the face consequently affected the automatic adjustment of the exposure times, which used the pixel values of this ROI as input. This loss led to underexposed or overexposed images, making HR assessments impossible. It also resulted in the inability to generate a continuous time series of facial regions, which made it impossible to determine the HR as well. In total, 12 out of the 36 recordings could not be analyzed and were removed from the assessment. In six of the failed recordings, the time-series length was insufficient for HR extraction, and in the other six, the faces were overexposed. An overview is shown in Table 2.

**Table 2.** Overview of valid frames and achieved mean facial exposure for each subject and recording phases.

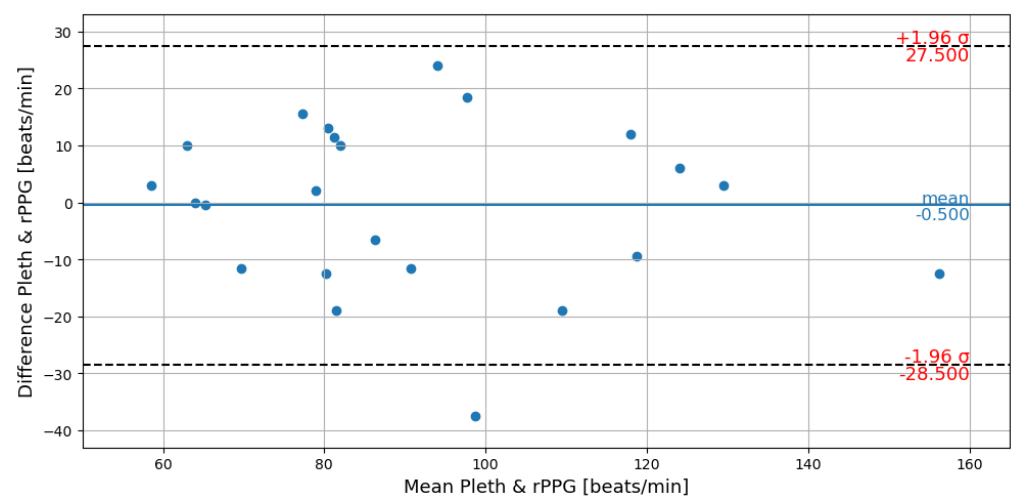
| Subject | Valid Frames | Phase I       |                  | Valid Frames | Phase II      |                  |
|---------|--------------|---------------|------------------|--------------|---------------|------------------|
|         |              | Mean Exposure | Observation      |              | Mean Exposure | Observation      |
| S01     | 423          | 208           |                  | 61           | 215           |                  |
| S02     | 306          | 185           |                  | -            | -             | face out of area |
| S03     | 212          | 253           |                  | 350          | 252           |                  |
| S04     | -            | -             | face out of area | -            | -             | face out of area |
| S05     | -            | -             | over exposure    | 163          | 229           |                  |
| S06     | -            | -             | over exposure    | -            | -             | over exposure    |
| S07     | 219          | 212           |                  | 260          | 214           |                  |
| S08     | 327          | 244           |                  | -            | -             | face out of area |
| S09     | 49           | 230           |                  | 390          | 114           |                  |
| S10     | 390          | 147           |                  | 390          | 230           |                  |
| S11     | 213          | 230           |                  | 394          | 230           |                  |
| S12     | 218          | 180           |                  | 280          | 232           |                  |
| S13     | 288          | 163           |                  | 208          | 252           |                  |
| S14     | -            | -             | over exposure    | -            | -             | face out of area |
| S15     | 279          | 253           |                  | 220          | 252           |                  |
| S16     | 170          | 225           |                  | 68           | 253           |                  |
| S17     | 365          | 228           |                  | -            | -             | face out of area |
| S18     | -            | -             | over exposure    | -            | -             | over exposure    |

Of the other 24 recordings, an accumulated time-series duration of well over 20 s was achieved in 9 sets. A total of 3 recordings were under 10 s, and the remain-

ing were in between 10 s and 20 s. Regarding the face-related exposure settings, in 14 of the 24 successful recordings, a mean facial pixel value  $\bar{v}$  within the target range of above 190 without signal clipping could be achieved. The remaining were split into 5 recordings that were underexposed ( $\bar{v} < 190$ ) and 5 recordings that were overexposed ( $\bar{v} > 250$ ).

### 3.3. Heart Rate Assessment

The overall performance of the proposed approach for all valid datasets is shown in Figure 7. The HR assessed by the UAS was compared against the HR-rate value provided by the reference pulsometer via a Bland Altman plot. The reference was calculated as the mean value of the HR values for the valid time points of the video recording. The combined value of the CHROM- and POS-assessed HRs produced an almost mean-value-free distribution of the determined values. Most values fell within the range of  $\pm 20$  bpm and were in the 95% confidence interval of  $-28.5$ – $27.5$  bpm.



**Figure 7.** Bland–Altman plot of reference Pleth and rPPG.

To investigate the influence of the system’s motion and exposure-time adjustments on the quality of the HR assessments, we split the 24 recordings into different subsets. Concerning system motion, we considered the total length of the time series extracted by the pipeline. The more frequently the face was lost due to the system motion, the shorter the total length. This way, the dataset was split into 11 recordings, representing high system motion when the length was shorter than 15 s (i.e., 225 samples) and 13 recordings of low system motion when the length was greater than 15 s. Regarding the exposure, the set was split into the 14 recordings where the facial color-channel values were within the desired range and the 10 recordings where they were not. The cross-relations for all color-mixing methods regarding the root-mean-squared errors (RMSEs) of all four sets, as well as the complete dataset, are shown in Table 3.

**Table 3.** RMSE bpm for the different subsets according to the recording conditions.

|                                | Recordings Outside Exposure Range |            | Recordings Inside Exposure Range |            | All Exposure Ranges  |            |
|--------------------------------|-----------------------------------|------------|----------------------------------|------------|----------------------|------------|
|                                | Number of Recordings              | RMSE (bpm) | Number of Recordings             | RMSE (bpm) | Number of Recordings | RMSE (bpm) |
| High Motion Recordings         | 5                                 | 19.3       | 6                                | 12.7       | 11                   | 16.0       |
| Low Motion Recordings          | 5                                 | 13.9       | 7                                | 11.4       | 12                   | 12.5       |
| High and Low Motion Recordings | 10                                | 16.8       | 13                               | 12.0       | 23                   | 14.3       |

The RMSE values showed that the approach benefited most from a more stable recording and good exposure (low motion, inside exposure range), resulting in an RMSE of 11.4, which was comparable to the other methods. The effects of image stability and exposure were comparable in our recordings. Optimal illumination (high and low motion, inside exposure range), as compared to a less optimal environment (high and low motion, outside exposure range), led to an improvement of the RMSE by 4.8 bpm, whereas a stable recording (low motion, all exposure ranges), as compared to an unstable one (high motion, all exposure ranges), led to an improvement of 3.5 bpm. Furthermore, the respective worst-case results (high motion, all exposure ranges) with 16.0 bpm vs. (high and low motion, outside exposure range) 16.8 bpm, and the best-case results (low motion, all exposure ranges) with 12.5 bpm vs. (high and low motion, inside exposure range) 12 bpm, were very similar. Both indicated that the influence of the exposure and the significant system motion had an equal effect on the quality of the acquired HR assessments.

#### 4. Discussion

Our research represented an initial exploration of the feasibility of acquiring vital parameters via drones under outdoor conditions. Furthermore, it evaluated the possibility of automatically assessing individual HRs in the context of MCI scenarios and using a specialized UAS. The data presented in the study were obtained through a feasibility investigation involving a small group of stationary and easily observable subjects. Therefore, the findings should be regarded as preliminary indications of the feasibility and serve only as a foundation for future research. To validate the findings and draw conclusive inferences, further studies with a larger sample of subjects representative of the population and a more extensive range of measured HRs are necessary, particularly under more realistic conditions.

Despite the challenging weather conditions during the study, we found that the developed tilt-wing UAS was capable of producing recordings appropriate for HR assessments. However, we also found that recording an individual's face for a sufficient length of time during powerful wind gusts was not feasible. In an actual application, this would have to be detected automatically, so the recording time could be extended accordingly.

Losing the facial region also affected the adaptive exposure. In the 14 flawless working recordings, the face was continuously within the frames during the adjustment. For the other recordings, the algorithm either did not adjust because no facial ROIs were recognized or arbitrary regions were incorrectly recognized as facial ROIs. The latter was mainly responsible for the highly overexposed recordings. If more powerful hardware becomes available, more robust, deep neural-network-based algorithms for facial detection could be used [33]. The fusion of several sensors, e.g., by using an additional thermal-imaging camera, could also ensure that only actual facial ROIs are detected.

Our results showed, additionally, that the general assessment of HRs via drones was possible. Based on the RMSE with slight system motion and sufficient exposure, the values achieved in our study were comparable with the results of other rPPG studies, which had been performed under much more controlled conditions. The 11.4 bpm RMSE we achieved was similar to the results of Nowara et al., who had achieved an RMSE of  $11.0 \pm 3.8$  bpm while driving under motion and using a similar approach [34]. However, the high-motion results showed much room for improvements in the proposed system. A more stable image, in which the face was consistently captured, and good illumination of the face could contribute equally to better results, lowering the RMSE of 19.3 bpm to 13.9 bpm and 12.7 bpm, respectively.

In contrast to a real-world scenario where higher HRs are caused by stress or medical conditions, the HRs after brief physical activity, such as those in our study, only remained stable for the limited duration of a few seconds before rapidly decreasing and returning to near-resting levels within one minute. This limited the evaluation of our method, as it relied on a stable HR over time and, thus, could result in less accurate detection of higher HRs. However, the data collected in our study, as displayed in Figure 7, did not demonstrate

significantly higher deviations at higher HRs. One potential explanation for this could be that the valid recording time periods in these cases were well below 20 s, thereby reducing the impact of this effect.

The achieved average RMSE of 14.3 bpm showed that our approach could not provide the precise HR measurements recorded in a clinical setting. However, it was suitable for general HR estimations during MCIs. In such scenarios, the exact HR is not the focus, but rather, determining whether the HR falls above or below specific thresholds is critical. Additionally, a manual pulse assessment, commonly used in emergency services, can also deviate by several bpm from the actual HR, particularly in overly stressful situations such as MCIs. When considering the worst-case results of almost 20 bpm RMSE, the pulse detection was well above the acceptable margin-of-error in triage. To address this limitation, future research could focus on obtaining more accurate results and developing an error measure that evaluates the quality of the respective measurements and notates poor measurements as unreliable.

Using a 3-axis gimbal integrated into the UAS could solve the most significant issues impeding the reliability of the system. One of the main reasons for the loss of the face region during recording and related overexposure was that the 2-axis gimbal could not compensate for the central disturbance axis. In addition, our results suggested that using a 3-axis gimbal could also increase the accuracy of the acquisition, as both the exposure and the available recording lengths would be improved. The algorithms for automatic exposure-time setting and robust facial recognition are typically integrated into the hardware of most modern consumer cameras. However, to our knowledge, no commercially available product combines a 3-axis gimbal with this camera technology that can also provide the images in raw format for evaluation. This currently makes the use of specialized drones with their own recording hardware a requirement.

Current research in rPPG is focused on training machine-learning-based methods for HR assessments. Previous research has shown that with more available training data, machine-learning-based methods have inevitably outperformed conventional approaches, which has also been shown in face-mesh-generation networks. Furthermore, due to our module-based approach, each step could be replaced by a machine-learning-based method to improve performance successively in the future. Improvements in recording the rPPG signal could also allow other blood-flow-related measurements to be assessed, such as blood pressure [35], which is also relevant for emergency medical services.

However, using visible light cameras, in general, and remote rPPGs, in particular, included their inherent limitations. It was essential that a person's face be recorded while unobscured, meaning that individuals not facing the drone could not be assessed at all. In addition, the facial ROIs should, if possible, not be covered. Although previous research has demonstrated that with facial hair and makeup, an rPPG measurement was still possible [36], it could be assumed that heavy occlusions by blood and dirt, which would be expected in MCI scenarios, could significantly impair the results of an HR assessment.

Additionally, the accuracy of HR detection using rPPGs was highly dependent on the amount of light reflected from the skin, which varies depending on the individual's skin type. Individuals with higher melanin content in their skin, which absorbs a greater proportion of environmental light than those with lower melanin content, tend to have less accurate rPPG measurements [25,37]. This could result in people with darker skin tones being disadvantaged during technically assisted triage. For these reasons, using a visible light camera alone may not be sufficient for reliable HR assessment during MCIs. Instead, it would be necessary to incorporate additional sensors, such as radar, which could record vital parameters and the visibility of certain facial regions, independently of skin color. However, it should be noted that radar systems are generally more expensive and more susceptible to system movements than cameras.

Theoretically, our approach could also be used on consumer-grade drones. They have the advantage that they are already available for most emergency services and first-responder units. Furthermore, these drones have the necessary hardware to perform all

the steps of our pipeline. However, the necessary data from the flight system sensors, the camera gimbal, and the raw camera data are typically unavailable to the end-users due to manufacturer restrictions, which limited our research to specialized flight systems.

Our special-purpose flight system had the advantage that its tilt-wing design and autonomous flight enabled it to be sent out simultaneously with the first-responders since, as compared to conventional drones, it could travel long distances. Thereby, it arrived on the scene before the emergency services. Along with the improved recording hardware and the possibility of assessing additional parameters relevant to triage, this could enable it to perform an automated triage in advance for emergency services.

In order to apply the presented approach in real-world applications, it is necessary to conduct further research on several factors that were intentionally disregarded in our study but are expected to affect the reliability during real MCIs. These include, but are not limited to, individuals in different poses, moving individuals, partially concealed individuals, covered faces, and individuals not facing the camera. Additionally, distinguishing between emergency responders and injured individuals is another critical factor that needs to be explored in an actual application.

## 5. Conclusions

This paper presented the first detailed study of the usage of rPPGs for HR assessments using a UAS in the context of MCI scenarios. We presented a detailed approach, incorporating both hardware and software concepts, for the most robust assessments possible. First, we detailed the hardware settings and necessary sensor information to design an HR assessment pipeline. Second, we showed that with this information, an assessment was possible within the scope of an estimation, despite the harsh environmental influences. The system could not evaluate one-third of the available records because the movement compensation required due to harsh weather; however, with moderate system motion, results were achieved that could compete with other approaches under controlled conditions. This marks the first step towards an automated triage system during MCIs. However, further research is needed to explore the impact of movement, body position, and the concealment of facial regions and to develop a method for identifying unreliable measurements.

**Author Contributions:** Conceptualization, L.M., A.F. and M.C.; methodology, L.M.; software, L.M.; validation, L.M. and C.B.P.; formal analysis, L.M.; investigation, L.M., I.B. and A.M.; resources, I.B., D.M. and M.C.; data curation, L.M.; writing—original draft preparation, L.M.; writing—review and editing, all authors; visualization, L.M.; supervision, A.F. and C.B.P.; project administration, I.B., A.M. and A.F.; funding acquisition, D.M., A.F. and M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Federal Ministry of Education and Research (BMBF), grant number 13N14772, under the FALKE (Flugsystem-Assistierte Leitung Komplexer Einsatzlage) project.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local ethics committee Klinisches Ethik Komitee (KEK) and the Center for Translational and Clinical Research (CTC-A), with ethical approval number EK 022/22 (protocol code CTC-A, #22-007, approved 2 June 2022).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations were used in this manuscript:

|      |                             |
|------|-----------------------------|
| UAS  | Unmanned Aerial System      |
| MCI  | Mass-Casualty Incident      |
| HR   | Heart Rate                  |
| rPPG | Remote Photoplethysmography |
| ROI  | Region-of-Interest          |
| GLM  | Generalized Linear Modeling |
| SNR  | Signal-to-Noise Ration      |
| RMSE | Root-Mean-Squared Error     |
| bpm  | beats-per-minute            |

## References

- Johnson, A.M.; Cunningham, C.J.; Arnold, E.; Rosamond, W.D.; Zègre-Hemsey, J.K. Impact of using drones in emergency medicine: What does the future hold? *Open Access Emerg. Med. OAEM* **2021**, *13*, 487. [\[CrossRef\]](#)
- FJ, E.R. Drones at the service of emergency responders: Rather more than more toys. *Emergencias Rev. Soc. Esp. Med. Emergencias* **2016**, *28*, 73–74.
- Konert, A.; Smereka, J.; Szarpak, L. The use of drones in emergency medicine: Practical and legal aspects. *Emerg. Med. Int.* **2019**, *2019*, 3589792. [\[CrossRef\]](#) [\[PubMed\]](#)
- Abrahamsen, H.B. A remotely piloted aircraft system in major incident management: Concept and pilot, feasibility study. *BMC Emerg. Med.* **2015**, *15*, 12. [\[CrossRef\]](#) [\[PubMed\]](#)
- Rosser, J.C.; Vignesh, V.; Terwilliger, B.A.; Parker, B.C. Surgical and Medical Applications of Drones: A Comprehensive Review. *J. Soc. Laparoendosc. Surg.* **2018**, *22*, e2018.00018. [\[CrossRef\]](#) [\[PubMed\]](#)
- Carrillo-Larco, R.; Moscoso-Porras, M.; Taype-Rondan, A.; Ruiz-Alejos, A.; Bernabe-Ortiz, A. The use of unmanned aerial vehicles for health purposes: A systematic review of experimental studies. *Glob. Heal. Epidemiol. Genom.* **2018**, *3*, e13. [\[CrossRef\]](#) [\[PubMed\]](#)
- Moeyersons, J.; Maenhaut, P.J.; Turck, F.D.; Volckaert, B. Aiding first incident responders using a decision support system based on live drone feeds. In Proceedings of the International Symposium on Knowledge and Systems Sciences, Tokyo, Japan, 25–27 November 2018; Springer: Singapore, 2018; pp. 87–100.
- Burke, C.; McWhirter, P.R.; Veitch-Michaelis, J.; McAree, O.; Pointon, H.A.; Wich, S.; Longmore, S. Requirements and limitations of thermal drones for effective search and rescue in marine and coastal areas. *Drones* **2019**, *3*, 78. [\[CrossRef\]](#)
- Chuang, C.C.; Rau, J.Y.; Lai, M.K.; Shih, C.L. Combining unmanned aerial vehicles, and internet protocol cameras to reconstruct 3-D disaster scenes during rescue operations. *Prehospital Emerg. Care* **2018**, *23*, 479–484. [\[CrossRef\]](#)
- Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote plethysmographic imaging using ambient light. *Opt. Express* **2008**, *16*, 21434–21445. [\[CrossRef\]](#)
- Yu, S.; Hu, S.; Azorin-Peris, V.; Chambers, J.A.; Zhu, Y.; Greenwald, S.E. Motion-compensated noncontact imaging photoplethysmography to monitor cardiorespiratory status during exercise. *J. Biomed. Opt.* **2011**, *16*, 077010.
- Aarts, L.A.; Jeanne, V.; Cleary, J.P.; Lieber, C.; Nelson, J.S.; Oetomo, S.B.; Verkruysse, W. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—A pilot study. *Early Hum. Dev.* **2013**, *89*, 943–948. [\[CrossRef\]](#) [\[PubMed\]](#)
- Villarroel, M.; Chaichulee, S.; Jorge, J.; Davis, S.; Green, G.; Arteta, C.; Zisserman, A.; McCormick, K.; Watkinson, P.; Tarassenko, L. Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. *NPJ Digit. Med.* **2019**, *2*, 128. [\[CrossRef\]](#)
- Jeyakumar, V.; Nirmala, K.; Sarate, S.G. Non-contact measurement system for COVID-19 vital signs to aid mass screening—An alternate approach. In *Cyber-Physical Systems*; Elsevier: Amsterdam, The Netherlands, 2022; pp. 75–92.
- Magdalena Nowara, E.; Marks, T.K.; Mansour, H.; Veeraraghavan, A. SparsePPG: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1272–1281.
- Takano, C.; Ohta, Y. Heart rate measurement based on a time-lapse image. *Med. Eng. Phys.* **2007**, *29*, 853–857. [\[CrossRef\]](#)
- Unakafov, A.M. Pulse rate estimation using imaging photoplethysmography: Generic framework and comparison of methods on a publicly available dataset. *Biomed. Phys. Eng. Express* **2018**, *4*, 045001. [\[CrossRef\]](#)
- Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1. [\[CrossRef\]](#)
- Lempe, G.; Zaunseder, S.; Wirthgen, T.; Zipser, S.; Malberg, H. ROI selection for remote photoplethysmography. In *Bildverarbeitung für die Medizin 2013*; Springer: Berlin/Heidelberg, Germany, 2013; pp. 99–103.
- Kumar, M.; Veeraraghavan, A.; Sabharwal, A. DistancePPG: Robust non-contact vital signs monitoring using a camera. *Biomed. Opt. Express* **2015**, *6*, 1565–1588. [\[CrossRef\]](#)
- Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. I.

22. Tomasi, C.; Kanade, T. Detection and tracking of point. *Int. J. Comput. Vis.* **1991**, *9*, 137–154. [[CrossRef](#)]
23. Lucas, B.D.; Kanade, T. An Iterative Image Registration Technique with an Application to Stereo Vision. In Proceedings of the IJCAI-81: 7th International Joint Conference on Artificial Intelligence, Vancouver, BC, Canada, 24–28 August 1981; Volume 81.
24. De Haan, G.; Jeanne, V. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [[CrossRef](#)] [[PubMed](#)]
25. Wang, W.; Den Brinker, A.C.; Stuijk, S.; De Haan, G. Algorithmic principles of remote PPG. *IEEE Trans. Biomed. Eng.* **2016**, *64*, 1479–1491. [[CrossRef](#)]
26. Wang, W.; den Brinker, A.C.; De Haan, G. Discriminative signatures for remote-PPG. *IEEE Trans. Biomed. Eng.* **2019**, *67*, 1462–1473. [[CrossRef](#)]
27. Rapczynski, M.; Werner, P.; Al-Hamadi, A. Effects of video encoding on camera-based heart rate estimation. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 3360–3370. [[CrossRef](#)] [[PubMed](#)]
28. Chen, J.H.; Tang, I.L.; Chang, C.H. Enhancing the detection rate of inclined faces. In Proceedings of the 2015 IEEE Trust-com/BigDataSE/ISPA, Washington, DC, USA, 20–22 August 2015; Volume 2, pp. 143–146.
29. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
30. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
31. Lugaresi, C.; Tang, J.; Nash, H.; McClanahan, C.; Uboweja, E.; Hays, M.; Zhang, F.; Chang, C.L.; Yong, M.G.; Lee, J.; et al. Mediapipe: A framework for building perception pipelines. *arXiv* **2019**, arXiv:1906.08172.
32. Kartynnik, Y.; Ablavatski, A.; Grishchenko, I.; Grundmann, M. Real-time facial surface geometry from monocular video on mobile GPUs. *arXiv* **2019**, arXiv:1907.06724.
33. Yan, M.; Zhao, M.; Xu, Z.; Zhang, Q.; Wang, G.; Su, Z. Vargfacenet: An efficient variable group convolutional neural network for lightweight facial recognition. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019.
34. Nowara, E.M.; Marks, T.K.; Mansour, H.; Veeraraghavan, A. Near-infrared imaging photoplethysmography during driving. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 3589–3600. [[CrossRef](#)]
35. Bajraktari, F.; Liu, J.; Pott, P.P. Methods of Contactless Blood Pressure Measurement. *Curr. Dir. Biomed. Eng.* **2022**, *8*, 439–442. [[CrossRef](#)]
36. Wang, W.; Shan, C. Impact of makeup on remote-ppg monitoring. *Biomed. Phys. Eng. Express* **2020**, *6*, 035004. [[CrossRef](#)] [[PubMed](#)]
37. Nowara, E.M.; McDuff, D.; Veeraraghavan, A. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 284–285.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

## Article

# Automated Unmanned Aerial System for Camera-Based Semi-Automatic Triage Categorization in Mass Casualty Incidents

Lucas Mösch <sup>1,\*</sup>, Diana Queirós Pokee <sup>1</sup>, Isabelle Barz <sup>2</sup>, Anna Müller <sup>1</sup>, Andreas Follmann <sup>1</sup>, Dieter Moormann <sup>2</sup>, Michael Czaplik <sup>1</sup> and Carina Barbosa Pereira <sup>1</sup>

<sup>1</sup> Department of Anesthesiology, Faculty of Medicine, RWTH Aachen University, 52074 Aachen, Germany

<sup>2</sup> Institute of Flight System Dynamics, RWTH Aachen University, 52062 Aachen, Germany

\* Correspondence: lmoesch@ukaachen.de

**Abstract:** Using drones to obtain vital signs during mass-casualty incidents can be extremely helpful for first responders. Thanks to technological advancements, vital parameters can now be remotely assessed rapidly and robustly. This motivates the development of an automated unmanned aerial system (UAS) for patient triage, combining methods for the automated detection of respiratory-related movements and automatic classification of body movements and body poses with an already published algorithm for drone-based heart rate estimation. A novel UAS-based triage algorithm using UAS-assessed vital parameters is proposed alongside a robust UAS-based respiratory rate assessment and pose classification algorithm. A pilot concept study involving 15 subjects and 30 vital sign measurements under outdoor conditions shows that with our approach, an overall triage classification accuracy of 89% and an  $F_1$  score of 0.94 can be achieved, demonstrating its basic feasibility.

**Keywords:** video; contactless; drones; extraction; mass-casualty incident; triage



**Citation:** Mösch, L.; Pokee, D.Q.; Barz, I.; Müller, A.; Follmann, A.; Moormann, D.; Czaplik, M.; Pereira, C.B. Automated Unmanned Aerial System for Camera-Based Semi-Automatic Triage Categorization in Mass Casualty Incidents. *Drones* **2024**, *8*, 589. <https://doi.org/10.3390/drones8100589>

Academic Editor: Bo Li

Received: 14 September 2024

Revised: 13 October 2024

Accepted: 15 October 2024

Published: 17 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Mass casualty incidents (MCIs) like train crashes or great floodings require rapid response and triage to save as many lives as possible. Traditional triage methods rely on first responders to assess patients and categorize them based on the severity of their injuries. The lack of routine caused by the infrequency of these incidents and the higher-than-usual psychological and emotional stress levels of the first responders often result in wrong assignments of the triage categories, leading to delays in treatment and potentially higher mortality rates [1].

In recent years, unmanned aerial systems (UASs) have emerged as promising emergency response and disaster management tools. UASs can cover long distances in a relatively short amount of time and reach places that are hardly accessible to rescuers [2,3] while bypassing environmental hazards [4,5]. Furthermore, their ability to be equipped with various sensors enables them to perform various tasks, such as searching, rescuing, surveillance, and delivering medical supplies.

To address the challenges of traditional triage methods in MCIs and leverage the benefits of UASs without occupying existing personnel, we propose a novel automatic UAS triage system. In contrast to the UAS based triage system proposed by Lu et al. [6], this UAS employs contactless vital sign assessment to provide triage category suggestions for emergency personnel, establishing a semi-automatic triage system based on aerial footage.

Current triage algorithms used by first responders in the field, such as the Simple Treatment and Rapid Transport (STaRT) [7], usually rely on detecting essential vital signs, including ambulation, responsiveness, respiratory rate (RR), pattern, pulse, open injuries, and bleeding. In previous research examining the use of UAS for triage, a trained operator communicated with the injured individuals via radio devices attached to a drone while

observing them with a thermal camera [8]. Although achieving similar categorization accuracy compared to standard procedures, this approach resulted in the pilot and operator being permanently occupied and performed significantly slower than current approaches. In addition, no assessment of vital parameters was used, resulting in them having to be ignored or roughly estimated. The Aerial Remote Triage System (ARTS) [9] represents an alternative approach that aims to minimize the criteria necessary for triage assessment by leveraging the capabilities of an integrated drone camera to capture visual data that human operators can analyze. Despite this approach, the feasibility of acquiring vital parameters remotely through camera footage has not been considered. Nevertheless, several methods are currently available to assess vital signs remotely from a great distance [10].

Imaging photoplethysmography (iPPG) or remote photoplethysmography (rPPG) has been proposed as a method for non-contact assessment of HR [11,12]. This technique, similar to contact PPG, uses a camera to measure the changes in light intensity reflected by the skin [13]. It has been successfully demonstrated by our previous research that this technology can be utilized on a UAS to assess heart rate for triage purposes [14].

Multiple approaches have been proposed for the non-contact RR assessment [15]. These approaches vary based on the type of sensor used and can be categorized as using thermal cameras, conventional cameras, depth cameras, laser vibrometers, or radars. Approaches utilizing visible-light or depth imaging detect a person's respiration-related movement by analyzing chest and shoulder movements over consecutive video frames using change of pixel intensity [16], optical flow tracking [17] or Eulerian video magnification [18]. Another approach used with visible-light cameras involves using convolutional neural networks (CNNs) to generate data-derived filter kernels to detect respiratory motion [19] or other kinds of neural networks that can estimate surrogates for respiratory rate [20]. In contrast, thermal imaging measures respiration-correlated temperature changes, typically around the nostrils [21]. Laser vibrometers and radars are also capable of working independently of ambient lighting. A proof-of-concept for RR assessment via a drone-mounted radar was presented by Jing et al. [22]. However, this approach did not compensate for high system motion caused by wind. It required a relatively close distance directly above a lying person, making it unsuitable for use in an MCI.

Likewise, using camera-based methods presents a similar challenge, as the drone's movement can result in artifacts that can affect the accuracy of the results required for triage. Hence, all previous approaches to RR assessment using drone footage required a relatively close distance to the subjects [23], or even landing on them [24], or it had no compensation for wind as they were tested only under an optimal environment [25]. Therefore, it is important to overcome the challenge of compensating for drone movement artifacts to ensure the results are accurate enough for triage.

During triage, it is crucial to evaluate a person's overall physical condition and mental state. However, in situations where direct evaluation techniques are not available, the assessment must rely on auxiliary indicators. Moreover, for the proposed UAS-based triage system, this evaluation must be conducted using the available sensors, such as using a brief video sequence captured by the drone's camera system. Assessing the ability to walk or move can provide a good overall estimate as an individual moving is not likely to be unconscious or significantly impeded. Furthermore, body position is an additional indicator, as lying individuals are more likely to be distressed than those sitting or standing. An initial attempt was made by our research group to assess the mental and general state of individuals from an elevated position using available techniques for recognizing their movements [26]. However, this needs to be expanded for use with drone camera footage and the ability to automatically assess body poses.

Therefore, the main challenges for creating a semi-automated UAS triage system include developing a novel triage algorithm that utilizes contactless vital parameter detection and an MCI concept that incorporates the system. Moreover, it is necessary to modify the existing algorithms for drone-based RR detection, which have been primarily tested under quasi-stationary conditions [23–25], to compensate for heavy UAS movements and ensure

reliable detection. Additionally, entirely new procedures for assessing the physical and mental state of individuals must be developed.

Accordingly, this paper introduces an MCI concept that leverages UAS capabilities for semi-automatic triage. It presents a newly developed triage algorithm based on the camera-based detection of vital parameters, including heart rate, breathing rate, body position, and movement, using a UAS. Additionally, it presents an approach for the robust remote assessment of RR using digital image stabilization and signal denoising based on UAS sensor data. This method is designed to address the challenges of moving systems during gusty weather, ensuring the robustness of the data for triage. The paper also proposes a novel approach for automatically classifying body pose and movement detection using ML-based keypoint detection and a subsequent classification network. Finally, it showcases the feasibility of the proposed system through a pilot concept study involving 15 subjects and 30 vital sign measurements under outdoor conditions.

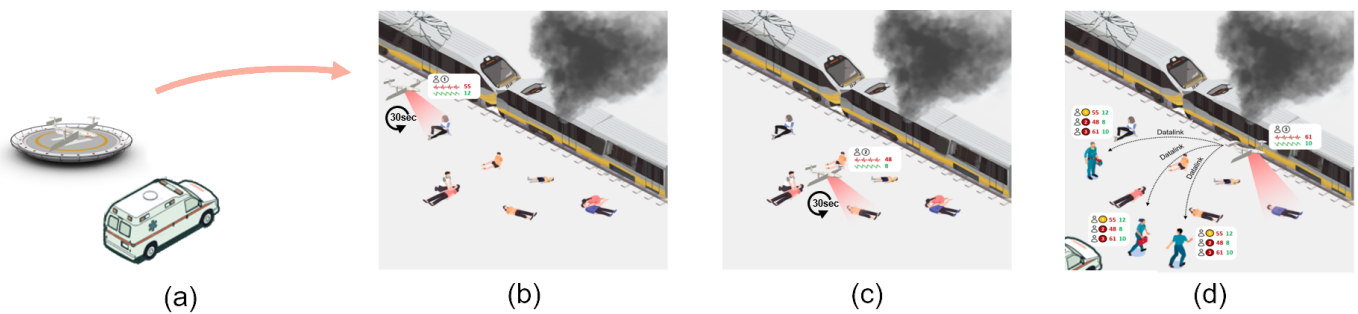
## 2. Materials and Methods

### 2.1. UAS Triage System

The UAS used for the proposed MCI concept and experimental evaluation is a tilt-wing UAS, which has been already presented in past publications [14]. It was developed by the Institute of Flight System Dynamics, RWTH Aachen University (Aachen, Germany), and the company flyXdrive GmbH (Aachen, Germany). The UAS is based on the “Neo” tilt-wing UAS manufactured by flyXdrive GmbH, provided during the “FALKE” research project, funded by the German Federal Ministry of Education and Research. This system can operate efficiently in both forward and hover flight modes, enabling vertical takeoff and landing in nearly any location. It can reach a top speed of up to 130 km/h during fixed-wing flight and maintain a hovering mode for several minutes. The UAS was further adapted to include gimbal-mounted sensors for the remote assessment of vital signs. As wind gusts can cause dynamic changes in position relative to the target and the UAS’s inherent movement, vibration, and shake during recordings, an electromechanical gimbal is needed to compensate for this. The gimbal utilizes a motor-driven cardanic suspension to maintain a constant line of sight relative to the target. The UAS was additionally equipped with a Nvidia Jetson Nano for on-board data processing.

#### 2.1.1. MCI Concept

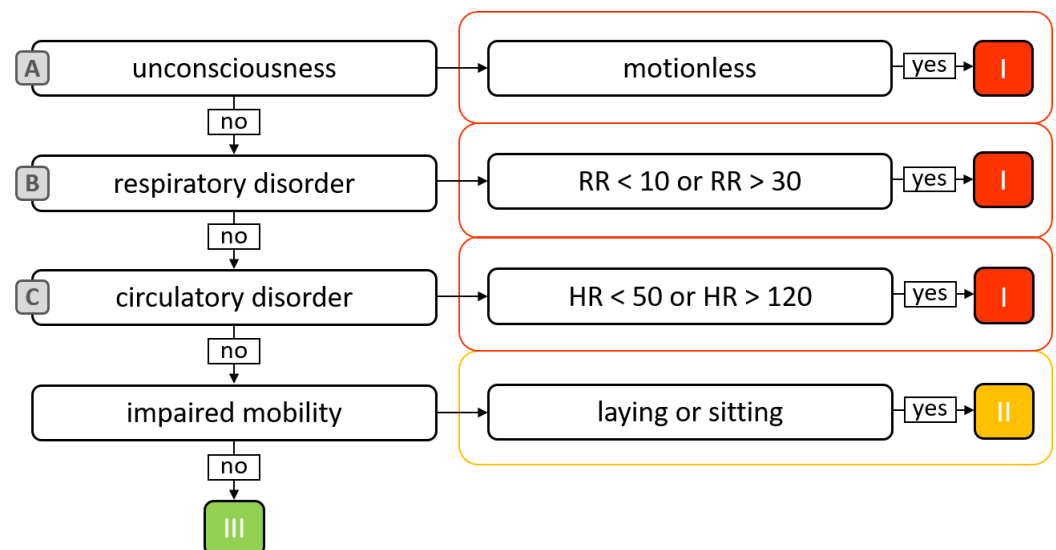
Our MCI concept provides that the UAS is already launched where it is stationed (Figure 1a). The UAS then navigates autonomously to a specified incident location and searches for injured people. For this, the UAS uses an integrated trajectory planner developed as part of the “FALKE” project [27]. The planner enables the UAS to autonomously approach any previously known location within its range. The UAS is not dependent on ground infrastructure and can even arrive at the operation area before the emergency personnel, saving crucial time due to its high speed in fixed-wing flight and its ability to bypass obstacles. Once it arrives at the incident location, the UAS commences its search for individuals who may be immobile and, therefore, likely injured. After these positions have been determined, each injured is successively approached for the non-contact assessment of vital signs (Figure 1b). For this purpose, the UAS approaches the patient and hovers at a position within recording distance for the assessment duration. The next injured person will then be approached (Figure 1c). Upon arrival of the emergency personnel, the collected data are processed and made available to the emergency services with a generated preliminary triage category and the recorded video footage (Figure 1d). Based on this information, the emergency physician in charge of the mission can decide whether to act according to the proposed triage or take different measures.



**Figure 1. Process of drone-based triage.** The UAS is deployed together with the emergency personnel (a). The UAS arrives at the scene and starts assessing the first immobile person (b). After the first assessment, the UAS approaches the next immobile person while processing the previous data (c). Upon arrival of the emergency personnel, the already processed data are transmitted wirelessly (d).

### 2.1.2. UAS Triage Algorithm

The algorithm for automated triage proposed in this paper, detailed in Figure 2, is based primarily on the two major MCI triage algorithms used in Germany: mSTaRT [28] and PRIOR [29]. The algorithms are designed to assist emergency responders in assessing the urgency of treatment needed. This is achieved by classifying the required treatment into five distinct categories. The first three categories are of particular importance: category I signifies the need for immediate treatment with the highest priority, category II represents urgent treatment that can be deferred and is of medium priority, while category III denotes that no acute treatment is required and is of low priority. To achieve this categorization, they rely primarily on the ABCDE schema by focussing on the examination of airway (A), breathing (B), circulation (C), mental disability (D) and environment (E) with the examination of **unconsciousness** (A), **respiratory disorder** (B), **circulatory disorder** (C), **bleeding** (C), **mental disorder** (D) and **impaired mobility**.



**Figure 2. Schema for UAS-based triage algorithm.** Illustrating the parameters and their ABCDE categories, along with the threshold values leading to the categorization into categories I, II, and III.

The goal in this work was to determine how the developed UAS could assess these characteristics using camera-based assessment methods. The detection of **unconsciousness** is covered through the recognition of body movement since a conscious person, especially when filmed by a UAS during an MCI, is likely to try to draw attention to themselves and thus is not motionless. The **respiratory disorder** is assessed by remotely detecting the RR. The remote assessment of HR covers **circulatory disorder**. To assess **impaired mobility**,

the person's pose is automatically evaluated. Assuming that a person sitting or lying down during an MCI is likely to be immobile, they are classified into three categories: lying, sitting, or standing.

## 2.2. Assessment of Body Pose and Movement

The assessment of body pose and movement relies on the machine-learning-based recognition of body key points (distinctive body landmarks like sternum, abdomen, and pelvic center and significant body joints) followed by the classification of body pose and the detection of limb movements. A multi-step approach is required to enable this assessment using image data from a UAS, comprising three major steps:

- (i) **Data acquisition and image pre-processing:** adjustments to the system's parameters prior to taking measurements and preparing the captured video frames;
- (ii) **Detection and tracking of body key points:** temporal tracking of relevant body key points and joints;
- (iii) **Classification of body pose and movement detection:** final step of pose classification and movement detection to be used for triage.

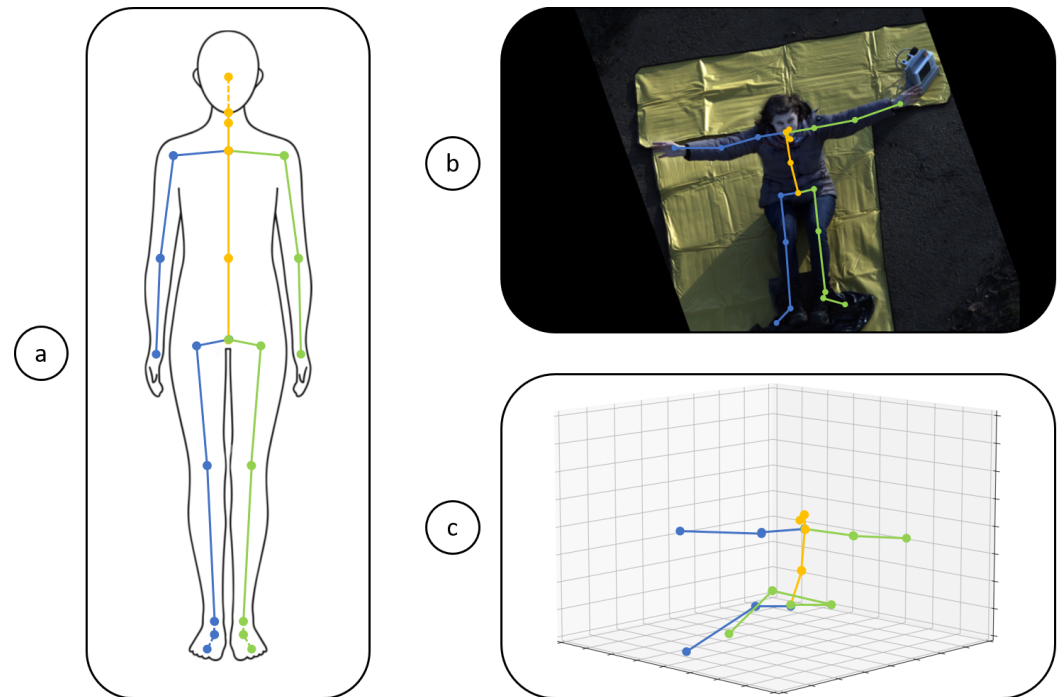
### 2.2.1. Data Acquisition and Image Pre-Processing

Before the triage evaluation, the UAS camera must first capture and process the video footage. For this purpose, the individual video frames denoted as  $I = [i_{t=0}, \dots, i_N]$ , are captured one by one in raw RGB format along with the timestamps  $t = [t'_0, \dots, t'_N]$ . Due to the limitation of using a two-axis gimbal that compensates only for the pitch and roll axes of the UAS, the captured images are not rotation-free in relation to the target and require correction. This correction can be accomplished by utilizing the sensor data from the gimbal, particularly by using the gimbal base vectors in conjunction with their respective timestamps, as presented in [14]. With this correction, a new set of rotation-free frames  $I_{rot} = [i_{rot,t=0}, \dots, i_{rot,N}]$  is generated for further processing.

### 2.2.2. Detection and Tracking of Body Key Points

After capturing and pre-processing the video footage, the next step is to locate the body key points within each frame of  $I_{rot}$ . For this purpose, the Efficient Video Pose Estimation via Neural Architecture Search (ViPNAS) approach proposed by Xu, Lumin, and Guan et al. [30] was employed. This multi-stage body key point detection, based on the MMPose toolbox of the OpenMMLab project, extracts 133 two-dimensional body key points from a given image or image sequence using multiple chained neural networks. Of these 133 points, 22 key points at relevant body joints were chosen for the pose classification, as illustrated in Figure 3a.

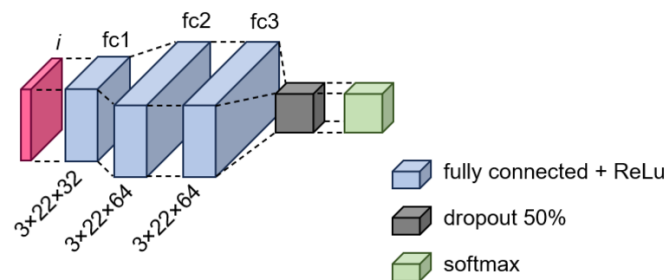
By applying this approach to each frame of  $I_{rot}$  individually, as exemplified in Figure 3b, a set of 2D body key points  $K_{2D} \in \mathbb{R}^{22 \times 2 \times N}$  is generated. In specific scenarios, mainly when dealing with individuals in upright or supine positions, these 2D key points can lead to ambiguities due to the near-identical spatial arrangements and interrelations. However, due to the inherent physiology of the human body, the arrangements between these key points are not arbitrary. The constraints of the human body's key points can be exploited to deduce the corresponding 3D positions from the available 2D data. To transform the 2D key points into 3D key points, the image-based 3D-lifter network provided by the MMPose Toolbox was used, which is based on the research conducted by Martinez [31]. Using this approach, a subset of the original 133 key points was transformed into 3D key points, and the corresponding 22 3D body key points  $K_{3D} \in \mathbb{R}^{22 \times 3 \times N}$  were generated from  $I_{rot}$ , as exemplified in Figure 3c.



**Figure 3. Three-dimensional body keypoint detection.** The relevant body key points on the human body for the left side (green), right side (blue), and center (yellow) (a), the detected points in the captured 2D video frames (b), and the reconstructed 3D points in a normalized 3D space (c) are depicted.

### 2.2.3. Classification of Body Pose and Movement Detection

In order to categorize the generated 3D body key points  $K_{3D}$ , a simple 2D fully connected neural network with rectified linear activation units (ReLU) was trained into the classes of laying (0), standing (1), and sitting (2). The specific architecture of this network is detailed in Figure 4.



**Figure 4.** Neural network structure of the pose classification network.

This network was then applied to each set of the 3D body key points corresponding to each frame of  $I_{rot}$ , resulting in the classification set  $C \in \{0, 1, 2\}^N$ . The final classification for the whole set was then obtained by applying the median to the whole set. The generated 3D key points were also used for motion detection. Eight 3D body joint angles  $\gamma_a - \gamma_h$  were defined to achieve this, as illustrated in Figure 5.

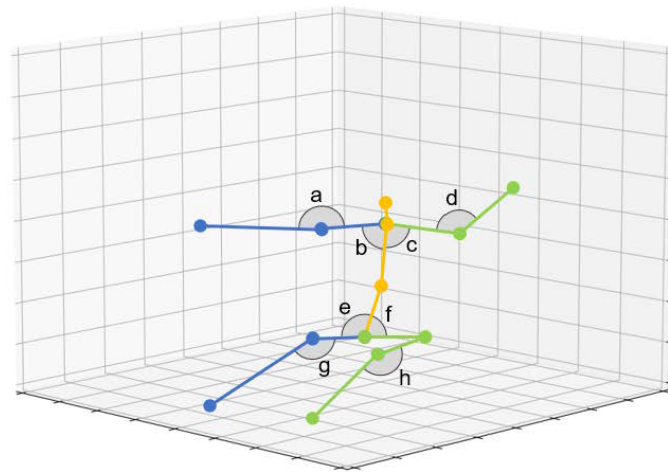
The angles between the joints offer a more robust approach than monitoring individual key points. This is because joint angles remain invariant to perspective changes and exhibit greater resilience to key point jitter. Movement detection is accomplished by analyzing the temporal changes of these angle values over each frame of  $I_{rot}$ :

$$\Gamma = \begin{bmatrix} \gamma_a \\ \vdots \\ \gamma_h \end{bmatrix} = \begin{bmatrix} \gamma_{a,t=0} & \cdots & \gamma_{a,t=N} \\ \vdots & \ddots & \vdots \\ \gamma_{h,t=0} & \cdots & \gamma_{h,t=N} \end{bmatrix} \quad (1)$$

If the standard deviation of any of the time series  $\gamma_a - \gamma_h$  exceeds the threshold  $t$ , the person is classified as moving and conscious:

$$d = \max \left( \begin{bmatrix} std(\gamma_a) \\ \vdots \\ std(\gamma_h) \end{bmatrix} \right) \begin{cases} d \geq t : & \text{movement} \\ d < t : & \text{no movement} \end{cases} \quad (2)$$

The threshold value of  $t = 0.1 \text{ Rad}$ , which corresponds roughly to  $6 \text{ Deg}$ , was chosen for the final classification.



**Figure 5. Body joint angles evaluated for motion detection.** Angles are evaluated symmetrically for the left side (green) and right side (blue), as well as for the center (yellow).

### 2.3. Assessment of RR

The assessment of the RR was made through the optical tracking of specific key points in the chest area of a person. This tracking was performed to capture respiration-related movements, allowing for the reconstruction of the respiratory curve. A multi-step approach was necessary to acquire curves suitable for RR analysis. This approach was designed to mitigate the motion artifacts introduced by the flight system and the additional movement of the individual and consists of four major steps:

- (i) **Data acquisition and image pre-processing:** System parameter adjustments made prior to conducting measurements and preparing the captured video frames;
- (ii) **Detection and tracking of chest region:** Temporal tracking of relevant chest region;
- (iii) **Detection and tracking of chest region key points:** Temporal tracking of relevant key points within the tracked chest region;
- (iv) **Extraction of respiratory curves:** Extraction and refinement of respiratory curves obtained through the temporal key point tracking;
- (v) **RR assessment:** Final step of assessing the RR for automated triage through spectral analysis.

#### 2.3.1. Data Acquisition and Image Pre-Processing

Including a larger image frame encompassing the whole chest area is advantageous when trying to capture chest movements effectively. Therefore, a different camera optic with an extended zoom factor has been employed for assessing the RR. However, the initial image processing procedures remain identical to those previously described in Section 2.2.1, resulting in a second set of  $I'_{rot}$ .

#### 2.3.2. Detection and Tracking of Chest Region

Within the rotated frames  $I'_{rot}$ , the region containing only the chest area shall be subsequently isolated and tracked across consecutive frames. In the first step, analogous to

the approach described in Section 2.2.2, the upper body's 2D key points  $K'_{2D} \in \mathbb{R}^{10 \times 2 \times N}$  are computed via ViPNAS for each frame. The chest region is subsequently defined as the square formed by the shoulder and hip points  $p$ , resulting in the bounding boxes:

$$BB = [bb_{t=0}, \dots, bb_N] = \begin{bmatrix} p_{1,t=0} & \cdots & p_{1,N} \\ p_{2,t=0} & \cdots & p_{2,N} \\ p_{3,t=0} & \cdots & p_{3,N} \\ p_{4,t=0} & \cdots & p_{4,N} \end{bmatrix} \quad (3)$$

However, these bounding boxes are not yet suitable for examining respiration, as they may still contain additional perspective distortions relative to each other, originating from uncompensated movements of the flight system. An affine image registration was conducted to mitigate these distortions without compromising the detection of respiration-related movements. A scale-invariant feature transform-based (SIFT) feature matching [32] was performed within the respective chest bounding boxes  $bb_n, bb_{n+1}$  for consecutive frames  $i_n, i_{n+1}$  to estimate the affine transformation parameters, resulting in the transformation  $T_n(x)$ . Subsequently, this transformation was applied to transform  $i_{n+1}$  to  $i_n$  and  $bb_{n+1}$  to  $bb_n$ . This procedure was iteratively applied to all frames of  $I'_{rot}$  and all bounding boxes of  $BB$ , thus generating a new set of images  $I_{tr}$  and boxes  $BB_{tr}$ . This approach is detailed in Algorithm 1.

---

**Algorithm 1** Consecutive Affine Image Warping

---

- 1: **let**  $i_0$  be the initial image frame
  - 2: **let**  $bb_0$  be the initial bounding box within  $i_0$
  - 3: **for**  $n = 0$  **to**  $N$ :
  - 4:      $T_n = SIFT(pixel(bb_n), pixel(bb_{n+1}))$
  - 5:      $i_{n+1} = T_n(i_{n+1})$
  - 6:      $bb_{n+1} = T_n(bb_{n+1})$
- 

When applied to real-world data, a potential challenge with this approach is the possibility of encountering situations where no bounding box or SIFT feature matching can be identified. This scenario is especially likely to occur when strong system movements result in the individual being partially or entirely outside the camera's field of view. In this case, a continuous image series of  $I_{tr}$  cannot be generated, and  $I_{tr}$  corresponds to individual connected segments  $I_{tr} = [I'_{tr}, I''_{tr}, \dots]$  and  $BB_{tr} = [BB'_{tr}, BB''_{tr}, \dots]$ , respectively. Therefore, only the longest continuous image sequence was selected for further analysis.

### 2.3.3. Extraction of Respiratory Curves

To extract the respiratory curves, capturing the respiration-related movements within  $BB_{tr}$ , is necessary. For this purpose, optical key point tracking, based on the principles of optical flow, was used. Initially, a grid of key points  $P_n$  was generated across the first bounding box's area, with

$$P_n = [p_1, \dots, p_M] = \begin{bmatrix} p_{x,1} \\ p_{y,1} \end{bmatrix} \dots \begin{bmatrix} p_{x,M} \\ p_{y,M} \end{bmatrix} \quad (4)$$

and with a horizontal and vertical spacing of 40 pixels. These key points' x and y coordinates were then tracked across all frames  $I_{tr}$  using optical flow tracking, generating a temporal set of 2D key point positions  $P = [P_{t=0}, \dots, P_N]$ . In order to generate the signal corresponding to the respiration for each key point's time series, the key points were orthogonally projected onto the corresponding longitudinal axis  $v$ , which was determined for each frame by calculating the best-fit axis between the points of the body center of  $K'_{2D}$ :

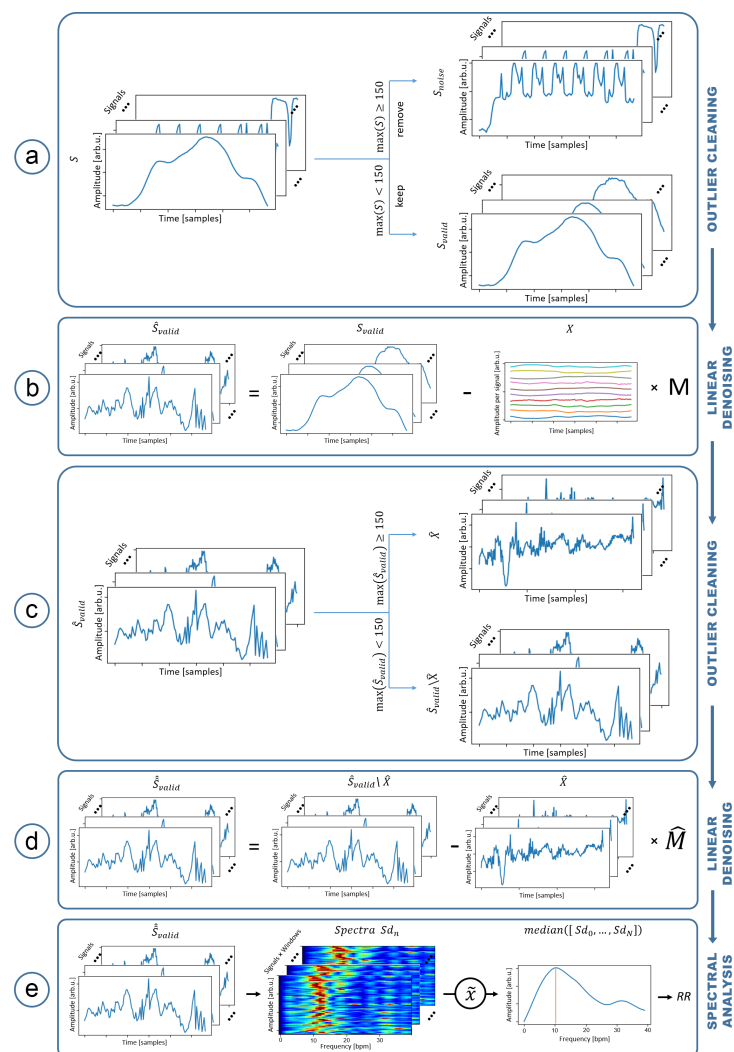
$$P_{Pr,n} = Pr_v(P_n) = [a_1, \dots, a_M] \cdot v = A_n \cdot v \quad (5)$$

After projection onto this axis, the respiratory curves  $S = [s_1, \dots, s_M]$  were measured as the deviation from the temporal mean position of  $A = [A_{t=0}, \dots, A_N]$ :

$$S = A - \text{mean}(A) = \begin{bmatrix} a'_{1,t=0} & \cdots & a'_{1,t=N} \\ \vdots & \ddots & \vdots \\ a'_{M,t=0} & \cdots & a'_{M,t=N} \end{bmatrix} = \begin{bmatrix} s_1 \\ \vdots \\ s_M \end{bmatrix} \quad (6)$$

### 2.3.4. Respiratory Rate Assessment

The whole process of RR assessment from the chest signals is shown in Figure 6. When determining a valid RR, the first step involves the removal of signals in  $S$  that are heavily corrupted, rendering them unsuitable for RR calculation (Figure 6a).



**Figure 6.** Overview of the RR estimation process. First signal separation (a). First denoising of  $S_{valid}$  with GPS and accelerometer data  $X$  (colored) (b). Second signal separation with  $\hat{S}_{valid}$  (c). Second denoising with  $\hat{S}_{valid} \setminus \hat{X}$  and separated signals  $\hat{X}$  (d) and RR estimation by frequency analysis using power density spectra  $Sd_n$  (red high, blue low) over sliding windows (e).

For this purpose, the signals in  $S$  were separated based on their maximum amplitude into two categories:  $S_{valid}$  with  $\max(S) \leq 150$ , representing the valid signals, and  $S_{noise}$  with  $\max(S) > 150$ , which is indicative of artifacts. In the second step (Figure 6b), linear denoising based on generalized linear modeling (GLM) was used on  $S_{valid}$ . For this,  $S_{valid}$

was modeled as a composition of signals from known sources of noise  $X$ , the noise mixing matrix  $M$ , the actual respiratory-related signals  $Y$ , and the residual noise  $E$ :

$$S_{valid} = X \cdot M + Y + E. \quad (7)$$

The objective was to subtract known noise signals  $X$  from  $S_{valid}$  according to their contributions  $M$ . To achieve this, the contributions  $M$  were determined through a least-squares fit for over-determined systems. Subsequently, the denoised signals  $\hat{S}_{valid}$  were obtained by subtracting  $X \cdot M$ :

$$\hat{S}_{valid} = S_{valid} - X \cdot M = Y + E \quad (8)$$

With this type of denoising, it is crucial to only use known noise signals that do not correlate with respiration to ensure that respiratory-related signals are not inadvertently removed from  $S_{valid}$ .

In our approach, the primary sources of noise are the artifacts caused by movements of the flight system. Therefore, signals that correlate with these movements but do not relate to respiration were required. For these signals, the data recorded during the flight through the camera gimbal were used. Specifically, the temporal deviation of the spatial vectors  $\vec{h}$ ,  $\vec{v}$  and  $\vec{s}$ , in addition to the calculated camera rotations  $\alpha$  of  $I_{rot}$ , were used:

$$X = \begin{bmatrix} h_{x,t=0} & \cdots & h_{x,t=N} \\ h_{y,t=0} & \cdots & h_{y,t=N} \\ h_{z,t=0} & \cdots & h_{z,t=N} \\ v_{x,t=0} & \cdots & v_{x,t=N} \\ v_{y,t=0} & \cdots & v_{y,t=N} \\ v_{z,t=0} & \cdots & v_{z,t=N} \\ s_{x,t=0} & \cdots & s_{x,t=N} \\ s_{y,t=0} & \cdots & s_{y,t=N} \\ s_{z,t=0} & \cdots & s_{z,t=N} \\ \alpha_{t=0} & \cdots & \alpha_{t=N} \end{bmatrix} \quad (9)$$

After denoising, it is possible that signals in  $\hat{S}_{valid}$ , which previously fell below the threshold of  $\max(S) > 150$ , are exceeding this threshold (Figure 6c). These signals, denoted as  $\hat{X} = \{\hat{s} \in \hat{S}_{valid} : \max(\hat{S}_{valid}) > 150\}$  were subsequently used in a second denoising pass (Figure 6d) alongside the remaining signals in  $\hat{S}_{valid}$ :

$$\hat{S}_{valid} = \hat{S}_{valid} \setminus \hat{X} - \hat{X} \cdot \hat{M} \quad (10)$$

resulting in the final denoised set of respiratory signals  $\hat{S}_{valid}$ .

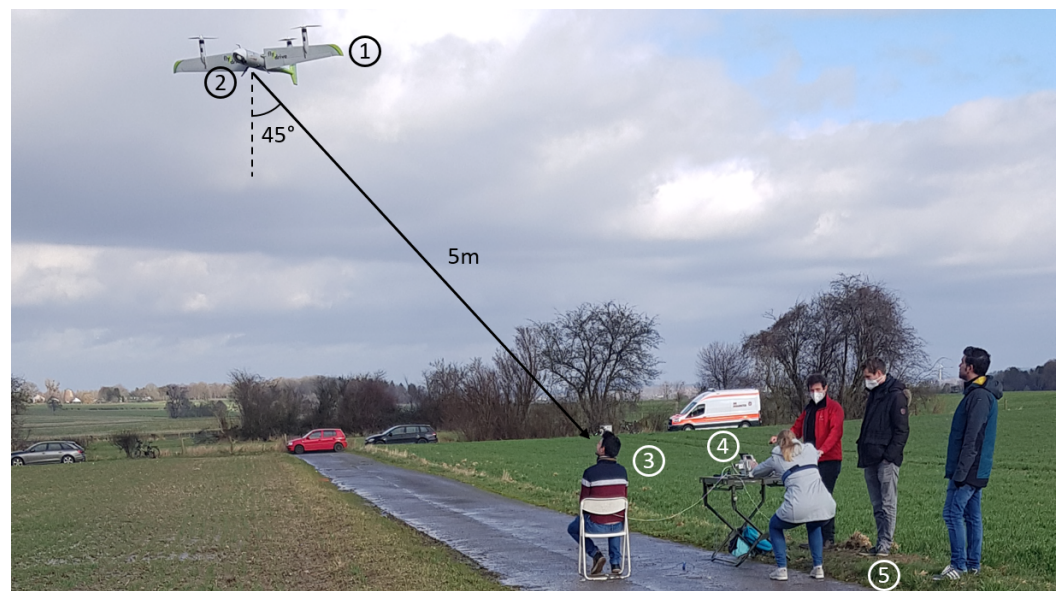
Most existing RR estimation approaches predominantly concentrate on continuous patient monitoring, emphasizing the detection of RR variations. However, in disaster medicine, this parameter's temporal variation may be less important than the immediate RR assessment for patient triage. Therefore, our approach focuses on obtaining a single reliable RR value. For RR assessment, it was assumed that RR changes only slightly over short time intervals. Thus, RR was considered constant within any chest region during the measurement. This assumption was used to compensate for short-term disruptions in the signals that the denoising process could not remove. All signals in  $\hat{S}$  were divided into overlapping segments of 100 samples using a sliding window with a stride size of one sample. For each signal window, the power-density spectrum  $Sd_n$  with a bandpass filter within the range of 5 to 40 breaths/min was calculated to determine the RR (Figure 6e). The final RR value was then determined as the maximum frequency of the median across all spectra:

$$RR = \operatorname{argmax}(\operatorname{median}([Sd_0, \dots, Sd_N])). \quad (11)$$

## 2.4. Experimental Setup

The dataset from a previous study [14] was repurposed to evaluate the newly developed automated triage system. In addition to the earlier publication on HR assessment, it is also used to evaluate the RR assessment. Furthermore, the acquired results of the HR assessment are reevaluated regarding our proposed approach to automated triage. Additionally, the first study was extended with whole-body recordings for evaluating the pose detection, involving 15 out of the 18 participants from the initial research using the same hardware in a different configuration.

The overall setup already published within the previous study [14] is shown in Figure 7. The goal was to record volunteers from a distance of five meters at a  $45^\circ$  horizontal angle. Since no live view was available, a fixed position was pre-measured for the volunteers, which the UAS targeted throughout the flight. Each volunteer sat in this position for two consecutive 30 s recordings. The volunteers were divided into two groups: Group One was tasked to simulate resting conditions, and Group Two should simulate abnormal respiratory rates by breathing very slowly or very quickly.



**Figure 7. Previous experimental setup:** UAS (1) equipped with gimbal and camera (2) is hovering in 5 m viewing distance and target engagement angle of  $45^\circ$  to a volunteering subject at a fixed position (3). The reference respiratory signal is recorded by a chest belt and stored on a computer (4). The rest of the respective subject group is preparing for their recording (5).

### 2.4.1. Flight System Sensors

The Neo tilt-wing UAS was fitted with a two-axis gimbal to balance the system motion during data recording. It could be controlled on the roll axis ( $-90^\circ$  to  $+90^\circ$ ) and pitch axis ( $-15^\circ$  to  $+60^\circ$ ) at an actuation speed of  $400^\circ/\text{s}$ . It was outfitted with an Allied Vision Mako G-234C RGB camera made by Allied Vision Technologies GmbH in Stadtroda, Germany. It utilizes a Sony IMX249 CMOS sensor from Sony Group Corporation in Tokyo, Japan. The selection of this camera was based on its compact size of 28 cm, lightweight at 80 g, and its capability to capture raw 10-bit Bayer frames at a maximum rate of 41.2 frames per second (fps), featuring a special resolution of up to 1936 by 1216 pixels. Furthermore, it uses a standard C-mount, allowing for the attachment of various optics, thus covering multiple distances.

### 2.4.2. Reference Telemetry

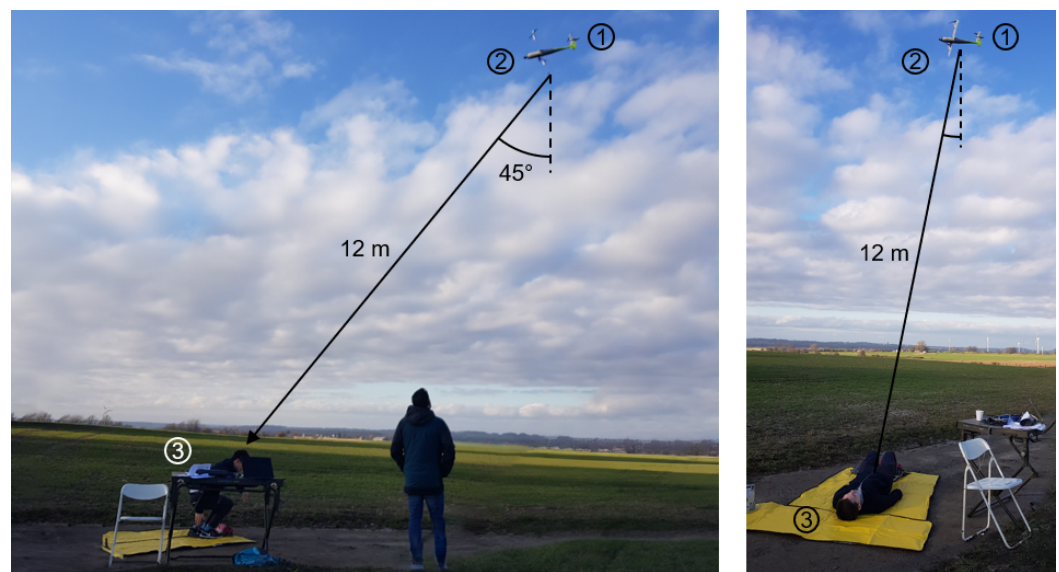
To obtain comparable reference values for the RR measurement, the volunteers were equipped with a respiratory chest belt (SleepSense PVDF Effort Sensor Kit (Adult)) by

S.L.P. Scientific Laboratory Products Inc., Elgin, IL, USA. This was in addition to the Philips MP2 used for HR reference assessment. The respiratory curves were captured using a NI USB-6000 ADC with a sampling rate of 30 kHz, by National Instruments, Austin, TX, USA. Respiratory belt signals and HR reference data were stored on a mobile computer and time-synchronized via WiFi and NTP with the UAS's onboard recording hardware to generate comparable references.

#### 2.4.3. Study Protocol for Pose Detection

To experimentally assess the efficacy of the newly developed triage algorithm, a cohort of 18 subjects underwent UAS recordings. The subjects were recruited from within the FALKE project members. The experiments were carried out with the cooperation of healthy volunteers who provided their written informed consent. Ethical approval for the study was granted by the ethics committee of the medical faculty at RWTH Aachen University in Germany (EK 022/22, protocol code CTC-A, #22-007, approved 2 June 2022). The volunteers included five females and ten males, aged between 20 and 36, predominantly of light skin color.

The overall setup for the evaluation of pose and movement classification is illustrated in Figure 8. The objective was to capture whole-body footage of volunteers from a horizontal viewing angle of  $45^\circ$ . As a Kowa 50 mm 1" 5MP C-Mount lens optic was used on the camera, this resulted in a targeted distance to the subject of 12 m. Two 30 s recording sessions, denoted as phase I and II, were conducted for each subject. Subjects were instructed to adopt a predetermined pose and perform either body movements or maintain stillness for each of the recordings. A comprehensive overview of the participants and their corresponding modalities is provided in Table 1.



**Figure 8. Experimental setup:** A UAS (1) fitted with a gimbal-mounted camera (2) is hovering at a view distance of 12 m and with an engagement angle of  $45^\circ$  to the subject (3), who is taking poses and/or movement.

To prevent participants from becoming cold and damp, particularly when assuming lying or sitting positions, a thermally insulating underlay was utilized during the recordings.

**Table 1. Subject modalities.** Overview over the modalities for movement and body pose for each subject and recording phase.

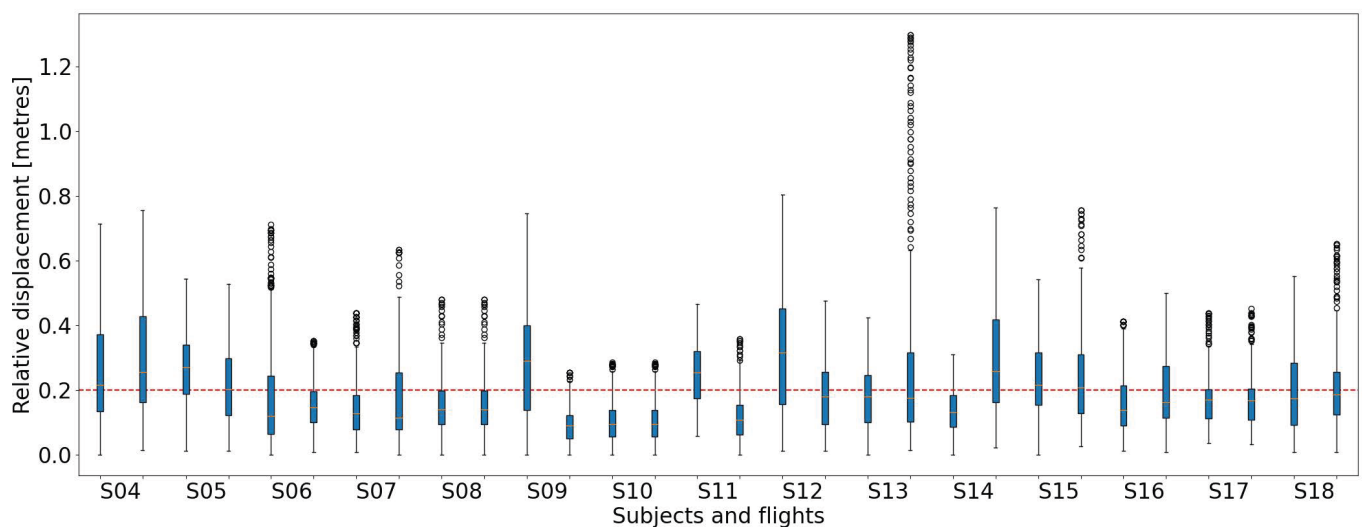
| Subject | Phase I  |         | Phase II |          |
|---------|----------|---------|----------|----------|
|         | Movement | Pose    | Movement | Pose     |
| S04     | Yes      | Sitting | No       | Sitting  |
| S05     | Yes      | Laying  | Yes      | Standing |
| S06     | No       | Sitting | Yes      | Sitting  |
| S07     | Yes      | Laying  | Yes      | Standing |
| S08     | No       | Sitting | Yes      | Standing |
| S09     | No       | Sitting | Yes      | Standing |
| S10     | Yes      | Laying  | Yes      | Laying   |
| S11     | No       | Laying  | Yes      | Laying   |
| S12     | Yes      | Laying  | Yes      | Sitting  |
| S13     | No       | Laying  | Yes      | Sitting  |
| S14     | No       | Sitting | Yes      | Standing |
| S15     | No       | Sitting | Yes      | Standing |
| S16     | No       | Laying  | Yes      | Sitting  |
| S17     | No       | Laying  | Yes      | Laying   |
| S18     | No       | Laying  | Yes      | Sitting  |

### 3. Results

#### 3.1. Recording Conditions

The recordings were made between 11 a.m. and 4 p.m. The outdoor conditions were demanding, with an average wind force of 6–7, equivalent to a wind speed of approximately 14 m/s, and strong wind gusts. Moreover, the weather shifted between sun, clouds, and light rain, leading to continually changing lighting conditions. The temperature was approximately 7 °C.

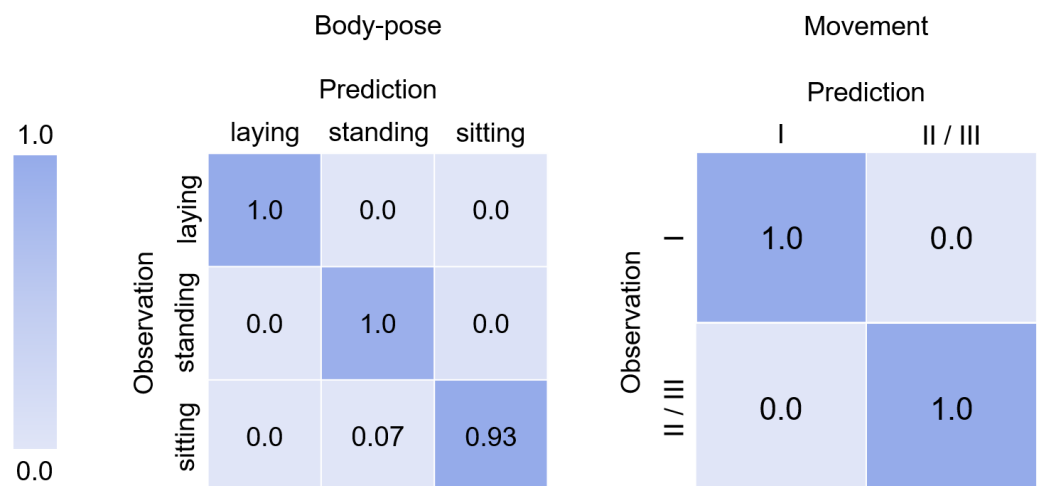
As a result of the demanding weather conditions, it was nearly impossible for the UAS to maintain a fixed position, resulting in significant system movement, as shown as relative displacement in Figure 9.

**Figure 9. UAS relative displacement:** Boxplot of the UAS displacement in m for every subject and flight.

Throughout almost all the flights, a general displacement of about 10 to 30 cm was observed with extreme cases where the UAS was pushed over 1.2 m away from its designated position. The average displacement across all flights was approximately 20 cm.

### 3.2. Assessment of Body Pose

To evaluate the body-pose assessment, a body-pose dataset was initially generated from valid frames of the aerial video footage, i.e., frames where the entire body is visible. This dataset assigns one of the classes, laying, standing, or sitting, to each recognized set of body key points in individual frames of the video recording, resulting in 11,255 keypoint–class pairs. Subsequently, the proposed neural network was trained using this dataset by splitting the dataset into training and test sets following the leaving-one-subject-out approach. Thus, the network was trained 15 times for each subject, with the data of 14 subjects, and evaluated on the remaining subject. The overall performance was then determined as the average of the best training performances for each respective subject, resulting in the cross-validation matrix shown in Figure 10 (left).



**Figure 10.** Cross-classification matrix for body-pose (left) and movement (right) assessment comparing the designated (observation) and assessed classification (prediction).

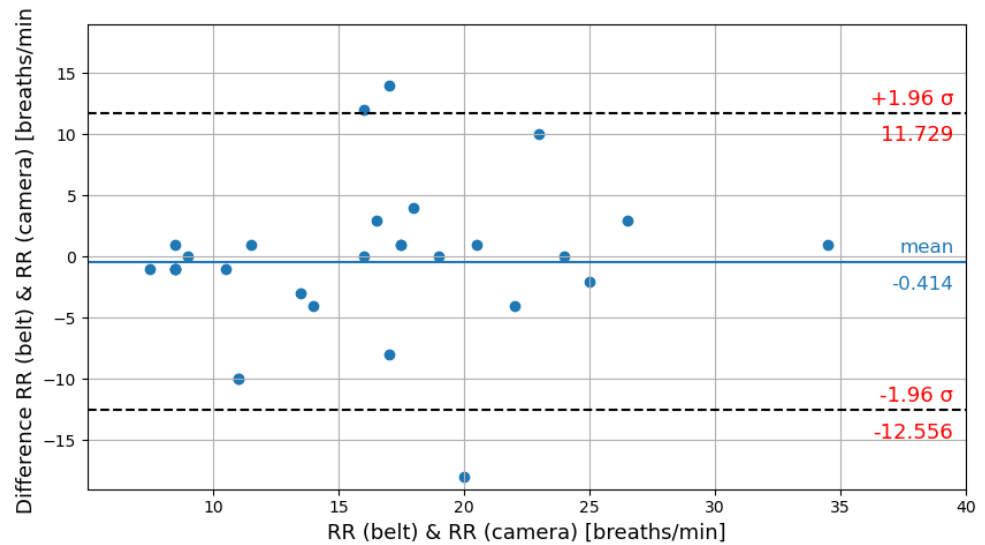
A 100% accuracy in classifying individuals as lying or standing could be achieved. Only once was a sitting subject misclassified as standing, resulting in a 93% accuracy for seated individuals. The normalized total classification accuracy was 97.6% with an  $F_1$  score of 0.98.

### 3.3. Assessment of Body Movement

The dataset used to assess body movement is derived from the dataset used for evaluating body-pose assessment. From the body-pose dataset, the longest contiguous sequence, wherein all body key points were observable, was selected for each subject and phase for evaluation. As shown in Figure 10 (right), our approach achieved perfect classification on this dataset, enabling a consistent differentiation between subjects in motion and those at rest.

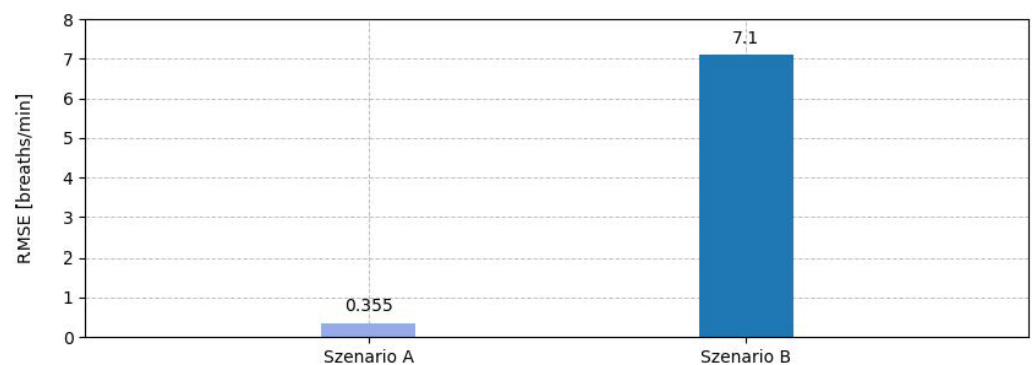
### 3.4. Assessment of Respiratory Rate

As described in [14], the subjects were instructed to simulate various physiological conditions, resulting in RR alterations. Accordingly, a dataset of upper-body videos and corresponding respiratory curves of 15 subjects for two phases was generated from the data to evaluate the RR assessment. The overall performance of the proposed approach for all valid datasets is shown in Figure 11. Through a Bland–Altman plot, the RR determined by the UAS approach was compared to the reference RR, which was derived by analyzing the frequency spectrum of the acquired respiratory curves. The reference RR was computed as the frequency corresponding to the spectral peak of the overall spectrum of the measurement. The majority of values fell within the range of  $\pm 5$  bpm and lie within the 95% confidence interval of  $-12.556$  to  $11.729$  breaths/min, which is primarily influenced by a few high outliers. The overall RMSE of the proposed approach resulted in 6.10 breaths/min.



**Figure 11.** Bland–Altman plot of reference RR and camera-based RR assessment.

In comparison to the results obtained by Al-Naji et al. [23], which were achieved at a significantly closer distance to the subjects and under quasi-stationary conditions, the proposed system exhibits a notably higher RMSE of 7.1 breaths/min, which is contrasted with the average RMSE of 0.355 breaths/min achieved by Al-Naji et al. [23], Figure 12.



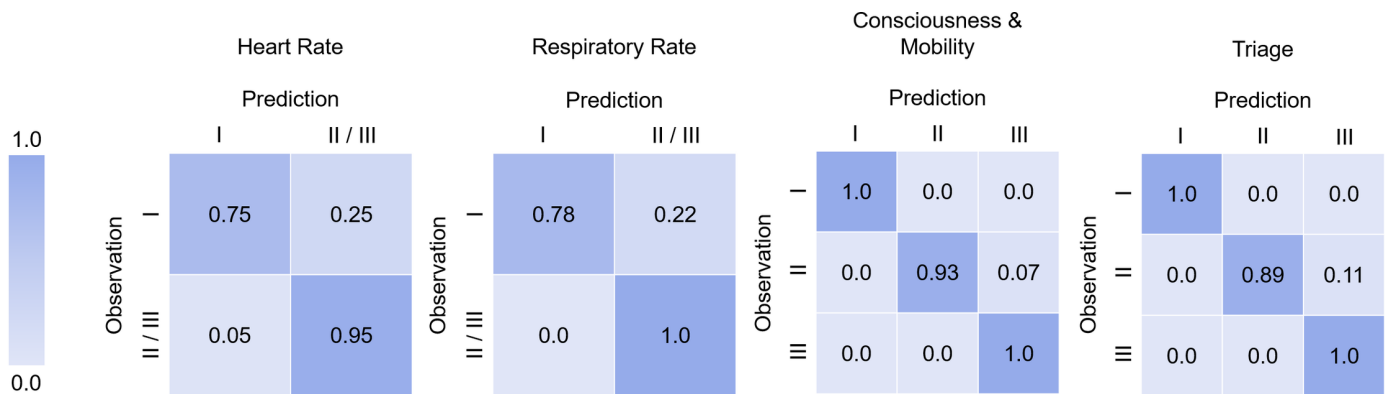
**Figure 12. Comparison of RMSE performance:** Comparison of a drone-based approach at a relatively close distance to the subject with minimal weather interference, as demonstrated by Al-Naji et al. [23] (Scenario A) versus the proposed system at greater distance with significant wind gusts (Scenario B).

### 3.5. Assessment of Heart Rate

The results for HR have been published in [14] and are accessible for review in this publication.

### 3.6. Triage Categorization

To evaluate the performance of the individual vital parameter assessments regarding their application for automated triage, each parameter was examined based on its direct impact on correctly altering the triage category according to the proposed schema. For HR and RR assessments, we specifically verified whether the evaluated values accurately led to the assignment of category I for our subjects. Regarding consciousness and mobility detection, it was verified whether participants were accurately classified into category II through body-pose detection and into category I through detection of immobility. The cross-classification matrices for the individual parameters are presented in Figure 13.



**Figure 13. Overview of triage cross-classification matrices.** The matrices depicting the individual impact of the vital parameters HR, RR, and consciousness and mobility on the proposed triage classification as well as the overall triage classification results.

Regarding HR, it shows that for 75% of individuals with a critical HR, the detection was accurate, leading to a correct assignment to category I. However, for the remaining 25%, the critical HR was not correctly assessed, resulting in no change in category. On the other hand, for 95% of participants with a non-critical heart rate, the identification was correct, and their category remained unchanged. For the remaining 5% with a non-critical heart rate, a critical heart rate was mistakenly identified, leading to an incorrect assignment to category I.

The results for RR were similar. In 78% of subjects with a critical RR, the detection was accurate, leading to their category I assignment. However, in 22% of subjects with a critical RR, correct detection did not occur, resulting in their non-assignment in category I. In contrast, all subjects with a normal RR were accurately identified, and there were no misclassifications. This results in a normalized overall accuracy of the approach of 89% and an  $F_1$  score of 0.94. The triage categorization based on consciousness and movement detection successfully identified all motionless subjects, thus correctly assigning them to category I. Among subjects with impaired mobility, 93% were accurately assessed and categorized as category II, while 7% were mistakenly identified as mobile and were not assigned to category II. On the other hand, all moving subjects without impairment were accurately identified, and the triage category remained unchanged. When combining the results of the individual parameters for all our study subjects, this results in the overall cross-classification matrix which is depicted in Figure 13.

#### 4. Discussion

Our study presents a preliminary investigation into the feasibility of semi-automatic triage using an automated UAS under outdoor conditions. Moreover, the feasibility of automatically assessing vital signs in MCI scenarios and implementing an adapted triage scheme was evaluated. The data for the study were gathered from a small group of relatively easily observable subjects at fixed positions, making the findings preliminary in nature. They serve only as a foundation for future research and should not be considered conclusive. More studies with a more prominent and representative subject sample and a wider range of simulated conditions are still necessary to validate the findings and draw conclusive inferences.

With its tilt-wing design and autonomous capabilities, our specialized flight system allows it to be deployed simultaneously with the emergency services. Unlike conventional drones, it covers longer distances and can reach the scene ahead of emergency personnel. With its integrated recording equipment and ability to evaluate various triage-related factors, it provides a preliminary automated triage before emergency personnel arrive.

Despite challenging weather conditions, the developed tilt-wing UAS produced recordings suitable for automated triage vital parameter assessment. This holds particularly true for the first-of-its-kind classification of body pose and movement, which allowed for an

almost perfect classification despite gusty winds during the flights. However, it should be noted that the provided thermally insulating underlay contributed to a higher contrast between individuals and the ground. In real-life scenarios, the achieved accuracy may not be easily reproducible.

The assessment of RR from a UAS has been demonstrated to be generally feasible. The achieved RMSE of 7.1 breaths/min is much larger as in previous studies conducted under highly controlled system conditions, e.g., the one by Janssen et al. [17] and is not yet suited to be used in the sense of clinical monitoring. Also, compared to the results by Al-Naji et al. [23], who could achieve an average RMSE of 0.355, the achieved RMSE of 7.1 breaths/min was much larger. This finding underscores that the error remains substantial despite the rigorous movement compensation, highlighting the considerable challenges associated with measuring respiratory rates under adverse weather conditions, particularly in the presence of gusty winds. However, given that the approach used by Al-Naji et al. [23] lacks motion compensation and was tested only under quasi-stationary conditions, it is expected to yield significantly worse results under the same conditions as those presented in our study. Nonetheless, it could be shown that our method is suitable for RR assessment within the context of triage, especially during gusty weather, where a correct triage classification rate regarding the influence of RR on the algorithm of 89% could be achieved. However, concerning the cross-classification rates of the approach, it reveals a high specificity and the requirement for further sensitivity enhancement for active usage.

This applies similarly to the performance of our approach for HR assessment. The evaluation within the triage algorithm reveals a normalized accuracy of 85% as well as demonstrating high specificity and indicating room for improvement in sensitivity in the cross-classification. In both HR and RR assessments, the primary reason for misclassifications are the high outliers during measurements, caused by harmonic noise, most likely as a result of the system motion. Such outliers inevitably impact the accuracy of triage classification. Therefore, in addition to opting for a more precise three-axis gimbal to reduce the impact of system motion, it is essential to develop a “goodness” metric that can identify and eliminate these outliers from the triage process.

The 93% overall accuracy achieved in the triage classification shows the promising performance of our approach. This is significantly more accurate compared to research studies that evaluated the effectiveness of human first responders. For instance, according to a study by Lee et al. [33] with briefly trained first responders on how to apply the SALT algorithm, humans scored between 57.8% and 87%, depending on their first responder background. Additionally, during the explicit simulation of an MCI, also by Lee et al. [33], the highest accuracy achieved by paramedic first responders was 79.9%. Our results even slightly outperform experienced first responders equipped with specialized medical equipment during a simulated MCI in a study by Follmann et al. [1] where an accuracy of 58.0% was achieved.

By capturing vital parameters, our approach is potentially more accurate in determining triage categories than Lu et al.’s approach [6], which relies solely on the classification of body positions. Furthermore, our method aligns more closely with established triage procedures, leading to higher acceptance by authorities and users.

However, the high overall accuracy of our approach is greatly influenced by the study subjects’ combination of HR, RR, and the designated poses and movements. When examining individual parameter performance, achieving a worst-case performance of only 75% is theoretically possible when the HR is the only critical parameter, which is still comparable to human accuracy. Unfortunately, there are little to no data on the distribution of conditions and vital parameters for real MCIs, as documenting scientific data in such situations could cost lives. Our current UAS-based triage system is designed for 30 s records of a person, which is twice the time taken by an experienced paramedic, who typically spends around 15 s [34]. However, a comparative or even faster average triage time can be achieved depending on the distance and obstacles between two individuals, as a UAS can typically travel much faster.

The proposed approach focused mainly on the technical capabilities of combining technologies and methods in conjunction with a UAS under outdoor conditions. To implement the presented approach in real-world scenarios, it is crucial to consider various factors intentionally omitted in our study but anticipated to impact reliability in genuine MCIs. This includes partially obscured individuals, diverse ground surfaces, and soiled clothing and faces. Also, nighttime triage scenarios were not addressed, where capturing heart rate through cameras is likely unfeasible, and detecting body pose and RR requires specialized night vision or thermal imaging cameras. Additionally, the ability to differentiate between emergency responders and injured individuals at the scene is a crucial factor that demands further exploration in real applications. Furthermore, the issue of racial bias in the use of remote photoplethysmography for HR assessment persists and necessitates consideration in the ongoing development of a usable system. In future work, the number of parameters relevant for triage could also be increased by utilizing semantic segmentation techniques to automatically detect bone fractures or bleeding. Also, reliance on wireless communication channels in real-world scenarios exposes the system to various vulnerabilities, including eavesdropping, jamming, and unauthorized access [35]. Given the sensitive nature of medical data, it is particularly crucial to protect it from unauthorized access. Accordingly, implementing secure communication [36], similar to military applications, is essential to ensure data integrity.

## 5. Conclusions

Our study provides a preliminary assessment of the feasibility of using a UAS for semi-automatic triage under challenging outdoor conditions. This work specifically focuses on the potential for automated vital sign assessment and triage in MCIs, leveraging an adapted triage scheme. While our findings offer promising insights, they are based on a small, relatively easily observable group of subjects in controlled environments. Therefore, these results serve as a foundation for future research rather than conclusive evidence. Our tilt-wing UAS, with its autonomous capabilities and specialized design, was able to capture vital triage-related data even under challenging weather conditions. Despite system movement due to wind, the UAS demonstrated accurate body-pose classification and delivered an overall triage classification accuracy of 93%, outperforming human first responders in several comparative studies. However, real-world applicability remains to be further validated, particularly concerning the motion compensation needed to improve HR and respiratory rate RR assessments. The study demonstrated the feasibility of robust RR assessment using the UAS during gusty wind, but the relatively high RMSE of 7.1 breaths per minute shows that significant improvement is required for clinical-grade monitoring. While the results indicate that UAS-based triage could be faster and potentially more accurate than traditional methods, the high performance was influenced by controlled conditions. Future research should address more diverse and complex environments, including nighttime scenarios, partially obscured individuals, and varying ground surfaces. Additionally, ethical considerations such as racial bias in rPPG need further investigation to ensure equitable and reliable triage outcomes. In conclusion, this study highlights the technical viability of UAS-based automated triage but underscores the need for further testing under more realistic and diverse conditions. By integrating additional capabilities, such as detecting fractures or bleeding through semantic segmentation, the UAS could become a more comprehensive tool for emergency response in the future.

**Author Contributions:** Conceptualization, L.M., D.Q.P., A.F. and M.C.; methodology, L.M.; software, L.M.; validation, L.M. and C.B.P.; formal analysis, L.M.; investigation, L.M., D.Q.P., I.B. and A.M.; resources, I.B., D.M. and M.C.; data curation, L.M.; writing—original draft preparation, L.M.; writing—review and editing, all authors; visualization, L.M.; supervision, A.F. and C.B.P.; project administration, I.B., A.M. and A.F.; funding acquisition, D.M., A.F. and M.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Federal Ministry of Education and Research (BMBF), grant number 13N14772 and 13N14773, under the FALKE (Flugsystem-Assistierte Leitung Komplexer Einsatzlage) project.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the local ethics committee Klinisches Ethik Komitee (KEK) and the Center for Translational and Clinical Research (CTC-A), with ethical approval number EK 022/22 (protocol code CTC-A, #22-007, approved 2 June 2022).

**Informed Consent Statement:** Written informed consent was obtained from all subjects before they participated in the study.

**Data Availability Statement:** The dataset presented in this article is not fully available due to privacy concerns. The data cannot be anonymized since the participant's faces must be visible, and not all subjects consented for their data to be shared publicly. However, a minimal dataset from participants who consented to data sharing is available at: <https://doi.org/10.5281/zenodo.12771420> (accessed on 16 October 2024).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

|      |                                     |
|------|-------------------------------------|
| UAS  | unmanned aerial system              |
| MCI  | mass-casualty incident              |
| HR   | heart rate                          |
| RR   | respiratory rate                    |
| rPPG | remote photoplethysmography         |
| ROI  | Region of Interest                  |
| GLM  | Generalized Linear Modelling        |
| SNR  | Signal-to-Noise Ratio               |
| RMSE | Root-Mean-Square Error              |
| bpm  | beats per minute/breaths per minute |

## References

1. Follmann, A.; Ohligs, M.; Hochhausen, N.; Beckers, S.K.; Rossaint, R.; Czaplik, M. Technical support by smart glasses during a mass casualty incident: A randomized controlled simulation trial on technically assisted triage and telemedical app use in disaster medicine. *J. Med. Internet Res.* **2019**, *21*, e11939. [\[CrossRef\]](#)
2. Roig, F.J.E. Drones at the service of emergency responders: Rather more than more toys. *Emer. Rev. Soc. Esp. Med. Emerg.* **2016**, *28*, 73–74.
3. Konert, A.; Smereka, J.; Szarpak, L. The use of drones in emergency medicine: Practical and legal aspects. *Emerg. Med. Int.* **2019**, *2019*, 3589792. [\[CrossRef\]](#)
4. Abrahamsen, H.B. A remotely piloted aircraft system in major incident management: Concept and pilot, feasibility study. *BMC Emerg. Med.* **2015**, *15*, 12. [\[CrossRef\]](#)
5. Rosser, J.C., Jr.; Vignesh, V.; Terwilliger, B.A.; Parker, B.C. Surgical and medical applications of drones: A comprehensive review. *J. Soc. Laparoendosc. Surg.* **2018**, *22*, e2018. [\[CrossRef\]](#)
6. Lu, J.; Wang, X.; Chen, L.; Sun, X.; Li, R.; Zhong, W.; Fu, Y.; Yang, L.; Liu, W.; Han, W. Unmanned aerial vehicle based intelligent triage system in mass-casualty incidents using 5G and artificial intelligence. *World J. Emerg. Med.* **2023**, *14*, 273. [\[CrossRef\]](#)
7. Garner, A.; Lee, A.; Harrison, K.; Schultz, C.H. Comparative analysis of multiple-casualty incident triage algorithms. *Ann. Emerg. Med.* **2001**, *38*, 541–548. [\[CrossRef\]](#)
8. Jain, T.; Sibley, A.; Stryhn, H.; Hubloue, I. Comparison of unmanned aerial vehicle technology-assisted triage versus standard practice in triaging casualties by paramedic students in a mass-casualty incident scenario. *Prehospital Disaster Med.* **2018**, *33*, 375–380. [\[CrossRef\]](#)
9. Álvarez-García, C.; Cámara-Anguila, S.; López-Hens, J.M.; Granero-Moya, N.; López-Franco, M.D.; María-Comino-Sanz, I.; Sanz-Martos, S.; Pancorbo-Hidalgo, P.L. Development of the aerial remote triage system using drones in mass casualty scenarios: A survey of international experts. *PLoS ONE* **2021**, *16*, e0242947. [\[CrossRef\]](#)
10. Molinaro, N.; Schena, E.; Silvestri, S.; Bonotti, F.; Aguzzi, D.; Viola, E.; Buccolini, F.; Massaroni, C. Contactless vital signs monitoring from videos recorded with digital cameras: An overview. *Front. Physiol.* **2022**, *13*, 160. [\[CrossRef\]](#)
11. Takano, C.; Ohta, Y. Heart rate measurement based on a time-lapse image. *Med. Eng. Phys.* **2007**, *29*, 853–857. [\[CrossRef\]](#)

12. Verkruyse, W.; Svaasand, L.O.; Nelson, J.S. Remote plethysmographic imaging using ambient light. *Opt. Express* **2008**, *16*, 21434–21445. [\[CrossRef\]](#)
13. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* **2007**, *28*, R1. [\[CrossRef\]](#)
14. Mösch, L.; Barz, I.; Müller, A.; Pereira, C.B.; Moormann, D.; Czaplik, M.; Follmann, A. For Heart Rate Assessments from Drone Footage in Disaster Scenarios. *Bioengineering* **2023**, *10*, 336. [\[CrossRef\]](#)
15. Massaroni, C.; Nicolo, A.; Sacchetti, M.; Schena, E. Contactless methods for measuring respiratory rate: A review. *IEEE Sens. J.* **2020**, *21*, 12821–12839. [\[CrossRef\]](#)
16. Massaroni, C.; Schena, E.; Silvestri, S.; Taffoni, F.; Merone, M. Measurement system based on RGB camera signal for contactless breathing pattern and respiratory rate monitoring. In Proceedings of the 2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA), Rome, Italy, 11–13 June 2018; pp. 1–6.
17. Janssen, R.; Wang, W.; Moço, A.; De Haan, G. Video-based respiration monitoring with automatic region of interest detection. *Physiol. Meas.* **2015**, *37*, 100. [\[CrossRef\]](#)
18. Wu, H.Y.; Rubinstein, M.; Shih, E.; Guttag, J.; Durand, F.; Freeman, W. Eulerian video magnification for revealing subtle changes in the world. *ACM Trans. Graph.* **2012**, *31*, 65. [\[CrossRef\]](#)
19. Pediaditis, M.; Farmaki, C.; Schiza, S.; Tzanakis, N.; Galanakis, E.; Sakkalis, V. Contactless respiratory rate estimation from video in a real-life clinical environment using eulerian magnification and 3D CNNs. In Proceedings of the 2022 IEEE International Conference on Imaging Systems and Techniques (IST), Virtual, 21–23 June 2022; pp. 1–6.
20. Brieva, J.; Ponce, H.; Moya-Albor, E. Non-contact breathing rate estimation using machine learning with an optimized architecture. *Mathematics* **2023**, *11*, 645. [\[CrossRef\]](#)
21. Kwon, H.M.; Ikeda, K.; Kim, S.H.; Thiele, R.H. Non-contact thermography-based respiratory rate monitoring in a post-anesthetic care unit. *J. Clin. Monit. Comput.* **2021**, *35*, 1291–1297. [\[CrossRef\]](#)
22. Jing, Y.; Qi, F.; Yang, F.; Cao, Y.; Zhu, M.; Li, Z.; Lei, T.; Xia, J.; Wang, J.; Lu, G. Respiration detection of ground injured human target using UWB radar mounted on a hovering UAV. *Drones* **2022**, *6*, 235. [\[CrossRef\]](#)
23. Al-Naji, A.; Perera, A.G.; Chahl, J. Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle. *Biomed. Eng. Online* **2017**, *16*, 101. [\[CrossRef\]](#)
24. Saitoh, T.; Takahashi, Y.; Minami, H.; Nakashima, Y.; Aramaki, S.; Mihara, Y.; Iwakura, T.; Odagiri, K.; Maekawa, Y.; Yoshino, A. Real-time breath recognition by movies from a small drone landing on victim's bodies. *Sci. Rep.* **2021**, *11*, 5042.
25. Al-Naji, A.; Perera, A.G.; Mohammed, S.L.; Chahl, J. Life signs detector using a drone in disaster zones. *Remote Sens.* **2019**, *11*, 2441. [\[CrossRef\]](#)
26. Queirós Pokee, D.; Barbosa Pereira, C.; Mösch, L.; Follmann, A.; Czaplik, M. Consciousness Detection on Injured Simulated Patients Using Manual and Automatic Classification via Visible and Infrared Imaging. *Sensors* **2021**, *21*, 8455. [\[CrossRef\]](#)
27. Barz, I.; Danielmeier, L.; Seitz, S.D.; Moormann, D. *Kollisionsfreie Flugbahnen für Kippflügelflugzeuge auf Basis des Wavefront-Algorithmus unter Berücksichtigung Flugmechanischer Beschränkungen*; Universitätsbibliothek der RWTH Aachen: Aachen, Germany, 2022.
28. Kanz, K.G.; Hornburger, P.; Kay, M.; Mutschler, W.; Schäuble, W. mStART-Algorithmus für Sichtung, Behandlung und Transport bei einem Massenanfall von Verletzten. *Notf. Rettungsmed.* **2006**, *3*, 264–270. [\[CrossRef\]](#)
29. Bubser, F.; Callies, A.; Schreiber, J.; Grüneisen, U. PRIOR: Vorsichtungssystem für Rettungsassistenten und Notfallsanitäter. *Rettungsdienst* **2014**, *37*, 730–734.
30. Xu, L.; Guan, Y.; Jin, S.; Liu, W.; Qian, C.; Luo, P.; Ouyang, W.; Wang, X. Vipnas: Efficient video pose estimation via neural architecture search. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16072–16081.
31. Martinez, J.; Hossain, R.; Romero, J.; Little, J.J. A simple yet effective baseline for 3d human pose estimation. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2640–2649.
32. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [\[CrossRef\]](#)
33. Lee, C.W.; McLeod, S.L.; Peddle, M.B. First responder accuracy using SALT after brief initial training. *Prehospital Disaster Med.* **2015**, *30*, 447–451. [\[CrossRef\]](#)
34. Follmann, A.; Ruhl, A.; Gösch, M.; Felzen, M.; Rossaint, R.; Czaplik, M. Augmented reality for guideline presentation in medicine: Randomized crossover simulation trial for technically assisted decision-making. *JMIR Mhealth Uhealth* **2021**, *9*, e17472. [\[CrossRef\]](#)
35. Krichen, M.; Adoni, W.Y.H.; Mihoub, A.; Alzahrani, M.Y.; Nahhal, T. Security challenges for drone communications: Possible threats, attacks and countermeasures. In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 9–11 May 2022; pp. 184–189.
36. Ko, Y.; Kim, J.; Duguma, D.G.; Astillo, P.V.; You, I.; Pau, G. Drone secure communication protocol for future sensitive applications in military zone. *Sensors* **2021**, *21*, 2057. [\[CrossRef\]](#)

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

### **Erklärung § 5 Abs. 1 zur Datenaufbewahrung**

Hiermit erkläre ich, dass die dieser Dissertation zu Grunde liegenden Originaldaten

- bei mir Lucas Mösch, Theresienstr. 18, 52072 Aachen  
sowie
- im AcuteCare InnovationHub der Klinik für Anästhesiologie des Universitätsklinikums  
Aachen

hinterlegt sind.

## Erklärung gemäß § 5 Abs. (1) und (2), und § 11 Abs. (3) 12. der Promotionsordnung

Hiermit erkläre ich, **Lucas Mösch**, an Eides statt, dass ich den wesentlichen Anteil an der Publikation:

**Mösch, L;** Barz, I; Müller, A.; Pereira, C.B.; Moormann, D.; Czaplik, M.; Follmann, A.:  
For Heart Rate Assessments from Drone Footage in Disaster Scenarios; *Bioengineering*;  
2023, 10, 336 geleistet habe

Die Anteile an der Arbeit waren wie folgt:

|                                      | Lucas Mösch | Isabelle Barz | Anna Müller | Carina B. Pereira | Dieter Moormann | Michael Czaplik | Andreas Follmann | Summe (%) |
|--------------------------------------|-------------|---------------|-------------|-------------------|-----------------|-----------------|------------------|-----------|
| Konzeption                           | 60          |               |             |                   |                 | 20              | 20               | 100       |
| Methodik                             | 100         |               |             |                   |                 |                 |                  | 100       |
| Algorithmenentwicklung / Software    | 100         |               |             |                   |                 |                 |                  | 100       |
| Validierung                          | 50          |               |             | 50                |                 |                 |                  | 100       |
| Auswertung                           | 100         |               |             |                   |                 |                 |                  | 100       |
| Planung und Durchführung Experimente | 40          | 30            | 30          |                   |                 |                 |                  | 100       |
| Bereitstellung von Ressourcen        |             | 20            |             |                   | 40              | 40              |                  | 100       |
| Datenverwaltung                      | 100         |               |             |                   |                 |                 |                  | 100       |
| <b>Verfassung des Manuskripts</b>    | <b>100</b>  |               |             |                   |                 |                 |                  | 100       |
| Korrektur des Manuskripts            |             | 20            | 15          | 20                | 15              | 15              | 15               | 100       |
| Überwachung und Koordination         |             |               |             | 50                |                 |                 | 50               | 100       |

Aus diesem wesentlichen Anteil ergibt sich selbstverständlich die Stellung als Erstautor.

---

Lucas Mösch  
Doktorand / korrespondierender Autor

Als Doktorvater oder Seniorautor bestätige ich die Angaben von Lucas Mösch.

---

Prof. Dr. Dr. Michael Czaplik  
Doktorvater

---

Dr. Andreas Follmann  
Seniorautor

Ich schließe mich der Erklärung von Prof. Dr. Dr. Michael Czaplik als Koautor an

---

Dr. Isabelle Barz

---

Anna Müller

---

Dr. Carina B. Pereira

---

Prof. Dr. Dieter Moormann

## Erklärung gemäß § 5 Abs. (1) und (2), und § 11 Abs. (3) 12. der Promotionsordnung

Hiermit erkläre ich, **Lucas Mösch**, an Eides statt, dass ich den wesentlichen Anteil an der Publikation:

**Mösch, L;** Pokee, D.Q.; Barz, I.; Müller, A.; Follmann, A.; Moormann, D.; Czaplik, M.; Pereira, C.B.:

Automated Unmanned Aerial System for Camera-Based Semi-Automatic Triage Categorization in Mass Casualty Incidents; *Drones* 2024, 8, 589 geleistet habe.

Die Anteile an der Arbeit waren wie folgt:

|                                      | Lucas Mösch | Diana Q. Pokee | Isabelle Barz | Anna Müller | Carina B. Pereira | Dieter Moormann | Michael Czaplik | Andreas Follmann | Summe (%) |
|--------------------------------------|-------------|----------------|---------------|-------------|-------------------|-----------------|-----------------|------------------|-----------|
| Konzeption                           | 40          | 20             |               |             |                   |                 | 20              | 20               | 100       |
| Methodik                             | 100         |                |               |             |                   |                 |                 |                  | 100       |
| Algorithmenentwicklung / Software    | 100         |                |               |             |                   |                 |                 |                  | 100       |
| Validierung                          | 60          |                |               |             | 40                |                 |                 |                  | 100       |
| Auswertung                           | 100         |                |               |             |                   |                 |                 |                  | 100       |
| Planung und Durchführung Experimente | 25          | 25             | 25            | 25          |                   |                 |                 |                  | 100       |
| Bereitstellung von Ressourcen        |             |                | 20            |             |                   | 40              | 40              |                  | 100       |
| Datenverwaltung                      | 100         |                |               |             |                   |                 |                 |                  | 100       |
| <b>Verfassung des Manuskripts</b>    | <b>100</b>  |                |               |             |                   |                 |                 |                  | 100       |
| Korrektur des Manuskripts            |             | 10             | 20            | 20          | 20                | 10              | 10              | 10               | 100       |
| Überwachung und Koordination         |             |                |               |             | 50                |                 |                 | 50               | 100       |

Aus diesem wesentlichen Anteil ergibt sich selbstverständlich die Stellung als Erstautor.

---

Lucas Mösch  
Doktorand / korrespondierender Autor

Als Doktorvater oder Seniorautorin bestätige ich die Angaben von Lucas Mösch.

---

Prof. Dr. Dr. Michael Czaplik  
Doktorvater

---

Dr. Carina B. Pereira  
Seniorautorin

Ich schliesse mich der Erklärung von Prof. Dr. Dr. Michael Czaplik als Koautor an

---

Diana Q. Pokee

---

Dr. Isabelle Barz

---

Anna Müller

---

Dr. Andreas Follmann

---

Prof. Dr. Dieter Moormann