Semi-Automatic Temperature Analysis Based on Real-Time Hand Landmark Tracking in Infrared Videos

A model to support research into the potential of infrared thermography for leprosy diagnosis

I.A.S. Schemkes



Delft University of Technology

TU Delft

Semi-Automatic Temperature Analysis Based on Real-Time Hand Landmark Tracking in Infrared Videos

A model to support research into the potential of infrared thermography for leprosy diagnosis

by

I.A.S. Schemkes

For the degree of Master of Science in Biomedical Engineering at the Delft University of Technology,
to be defended publicly on Wednesday July 10, 2024 at 1:45 PM.

Student name:	Irene Schemkes	
Student number:	4557468	
Track:	Medical Phyiscs	
Project duration:	December, 2023 – July, 2024	
Thesis committee:	Prof. dr. J. Dankelman (advisor),	TU Delft
	Dr. A. J. Knulst (advisor),	INFs Green Pastures Hospital, TU Delft
	Dr. J. C. van Gemert,	TU Delft
	Dr. J. W. Brandsma,	INFs Green Pastures Hospital

This thesis is confidential and cannot be made public until July 10, 2026.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

Preface

About ten months ago, I started my graduation project as the final chapter of my studies. Having started a bachelor's in Clinical Technology, I switched to a bridging program in Biomedical Engineering, focusing less on clinical aspects and more on imaging and image processing within the Medical Physics track. I was thrilled to jump into a project involving infrared imaging after learning about the research from Arjan. It sounded like a perfect blend of scientific, concrete, beneficial, and socially relevant research. After writing a literature review on the topic during the first few months, I started executing the research in December.

Thanks to the collaboration between Arjan and Jenny, I was lucky to have quickly formed a pleasant team of supervisors. I really appreciate the help and advice you have given me throughout this journey. We met regularly to discuss progress, and I could always pop in to brainstorm ideas or ask for advice. Your insightful feedback kept me on my toes and improved the quality of my thesis. Additionally, thanks to Jan for joining the project and providing technical insights, serving as extra inspiration. I am excited that all three of you will be there on July 10th, marking the official end of this past year and my time studying in Delft.

Besides, I want to thank my friends and family for their support. Lisa started the journey in Delft with me during our bachelor's, and we spent a lot of time this past year as we were both busy finishing our studies. She was always there to lend a hand, even with the time she had to dedicate to her own research. My parents and brother Lennart always gave me lots of love and a boost when needed and helped me get through the final days of deadline work.

It feels strange to be closing this chapter as a student, but I am eager to experience what comes next. Thanks to all the extracurricular activities I have done alongside my studies, I am really looking forward to starting a ”grown-up” job soon. I'm curious to find out my next step after I have enjoyed my summer vacation.

I.A.S. Schemkes
Delft, July 2024

Abstract

Leprosy is a neglected tropical disease affecting people worldwide. Left untreated or starting therapy too late, the infection can cause permanent damage to the nerves, skin, eyes, and limbs of patients. While current standard diagnostic techniques generally detect the disease once symptoms have manifested, infrared thermography presents a potential tool for the early detection of leprosy-related neuropathy. A research initiative has been established to explore whether autonomic nerve impairment in the hands can be identified using infrared thermography, necessitating a new method for temperature evaluation. This work discusses the development of a semi-automatic temperature analysis method for twelve regions of interest on the hands. The approach utilizes real-time landmark tracking based on an existing machine-learning model, adapted to the context of the application. After applying a cold impulse and recording the response with an infrared camera, the rewarming temperature and recovery values are determined within the regions. Results presented in this report indicate no significant improvement compared to the existing manual method. The success of the model outcomes highly depends on the accuracy of the machine-learning-based hand landmark detection, resulting in insufficiently reliable outcomes. The underlying explanations and the potential for future improvements are further discussed in this thesis, considering the evident potential and benefits of the automated approach.

Contents

Preface i

Abstract ii

Nomenclature v

1 Introduction 1
- 1.1 Motivation . . . 1
  - 1.1.1 Leprosy . . . 1
  - 1.1.2 Diagnosis and treatment . . . 3
  - 1.1.3 Infrared as a potential tool . . . 3
  - 1.1.4 Research initiative . . . 3
  - 1.1.5 Manual temperature analysis . . . 4
- 1.2 Research objective . . . 4
  - 1.2.1 Automatic temperature analysis . . . 4
  - 1.2.2 Model requirements . . . 4
  - 1.2.3 Study approach . . . 4

2 Background 6
- 2.1 Related literature . . . 6
- 2.2 Manual temperature analysis . . . 6
  - 2.2.1 Regions of interest . . . 6
  - 2.2.2 Algorithmic design . . . 6

3 Method 8
- 3.1 Data acquisition . . . 8
  - 3.1.1 Experimental setup . . . 8
  - 3.1.2 Cold pressure test . . . 9
  - 3.1.3 Infrared camera . . . 9
- 3.2 Software selection . . . 9
- 3.3 Algorithmic design . . . 10
  - 3.3.1 Video import and trimming . . . 10
  - 3.3.2 Mediapipe hand landmark tracking . . . 11
  - 3.3.3 Region of interest selection . . . 13
  - 3.3.4 Coordinate filtering . . . 13
  - 3.3.5 Temperature extraction . . . 15
  - 3.3.6 Variables of interest . . . 16
  - 3.3.7 Performance report . . . 16
  - 3.3.8 Data export . . . 17
- 3.4 Executive file . . . 18
- 3.5 Validation . . . 18
  - 3.5.1 Output overview . . . 18
  - 3.5.2 Repeatability . . . 18
  - 3.5.3 Visual examination . . . 19
  - 3.5.4 Effect of coordinate filtering . . . 19
  - 3.5.5 Comparison with the manual approach . . . 20
  - 3.5.6 Compare experiment set-ups . . . 20

4 Results 22
- 4.1 Video details . . . 22
- 4.2 Model output . . . 22

4.2.1 Output overview . . . 22
4.2.2 Example videos . . . 24
4.3 Validation . . . 25
4.3.1 Repeatability . . . 25
4.3.2 Visual examination . . . 27
4.3.3 Effect of coordinate filtering . . . 28
4.3.4 Comparison to the manual method . . . 30
4.3.5 Experiment set-up comparison . . . 34

5 Discussion **37**
5.1 Key results . . . 37
5.2 Results analysis . . . 37
5.2.1 Output overview . . . 37
5.2.2 Repeatability . . . 38
5.2.3 Visual examination . . . 39
5.2.4 Effect of coordinate filtering . . . 39
5.2.5 Comparison to the manual method . . . 39
5.2.6 Experiment set-up comparison . . . 39
5.3 Model requirements . . . 40
5.4 Limitations . . . 40
5.4.1 MediaPipe hand landmark tracking . . . 40
5.4.2 Coordinate filtering . . . 41
5.5 Recommendations for future research . . . 42

6 Conclusion **43**

References **44**

A User guide: Experiment Protocol **47**

B User Guide: Semi-Automatic Temperature Analysis **50**

C In-Depth Results **52**

Nomenclature

Abbreviations

Abbreviation	Definition
3D	Three-Dimensional
ANI	Autonomic Nerve Impairment
ANS	Autonomic Nerve System
BI	Bacterial Index
CPT	Cold Pressure Test
GUI	Graphical User Interface
IQR	Interquartile Range
IR	Infrared
IRT	Infrared Thermography
LoA	Limits of Agreement
MB	Multibacillary
MDT	Multidrug Therapy
MFT	Monofilament Test
NaN	Not a Number
OCR	Optical Character Recognition
PC	Paucibacillary
RC	Repeatability Coefficient
SD	Standard Deviation
ST	Sensory Testing
TAIL	Thermographic Assessment of Autonomic Impairment in Leprosy
VMT	Voluntary Muscle Test
WHO	World Health Organization

1 Introduction

1.1. Motivation

1.1.1. Leprosy

Leprosy, also known as Hansen's disease, is a chronic condition caused by Mycobacterium leprae (M. leprae) and is defined as a neglected tropical disease. The highest occurrence of this infectious disability is in developing countries and tropical areas [1]. According to the World Health Organization (WHO), a total of 165.459 prevalent cases and 174.087 new cases were reported in 182 contributing countries in 2022. Figure 1.1 shows the global distribution of new leprosy cases in 2020. Most incidences were registered in Southeast Asia and Africa [1]. The countries Indonesia, India and Brazil suffered the greatest number of cases, with more than 10.000 reported in 2020 [2]. M. leprae is transmitted by droplets from the mouth and nose of infected people. There is no risk of infection with brief contact, but only after months of close exposure [3].

Leprosy is a neuropathy and presents itself with the destruction of cutaneous nerve endings and trunks of peripheral nerves predominately [4]. This infectious damage negatively influences the motor, autonomic and sensory functioning of the neural system, with the latter causing the most clinically visible and common symptoms [5]. Usually, skin lesions are the earliest observable indication of the disease; ulcerations and deformities potentially follow [1, 6]. Left untreated, the nerves, skin, eyes and limbs of leprosy patients are at risk of being damaged permanently [4]. Apart from physical complaints, discrimination and stigmatisation are problems leprosy patients face [3].

The immune responses caused by M. leprae can be categorized into two main groups. Type I reactions result from the hypersensitivity and cellular response to mycobacterial antigens and can be reversed when treated timely [7]. Common manifestations are edema (swelling), hyperesthesia (increased sensitivity to sensory stimuli), ulceration, erythema (redness of the skin), neuritis and subsequent scaling (skin patches) [8, 9]. Type II or erythema nodosum leprosum comprises more severe, acute inflammatory reactions of the nerves, blood vessels, lymphatic vessels and the skin due to an acute immune complex. The patient's complete system is affected, causing symptoms such as fever [10, 11]. A frequently occurring symptom of type II leprosy is the emergence of painful erythematous subcutaneous nodules, red nodules beneath the skin, which have the potential to develop ulcers [12]. Lucio's phenomenon cannot be classified under type I or II reactions and is a very exceptional third-response class when the disease is persistent and untreated. This reaction variant presents itself with necrotic ulcers, among other things, and patients may die from superinfection and sepsis [1, 13]. Figure 1.2 shows visible manifestations of the different reaction types.

The WHO introduced a simplified classification system based on the bacterial index (BI), i.e. leprosy bacilli density retrieved from the slit-skin test. The paucibacillary (PB) class denotes individuals with a BI <2, typically involving one to five skin lesions and/or only one affected nerve trunk. Multibacillary (MB) cases with a BI ≥2 are regularly characterized by more than five skin lesions and/or multiple affected nerve trunks [1, 3, 14]. Since treatment duration depends on the distinction made between PB and MB cases, adequate classification is important.

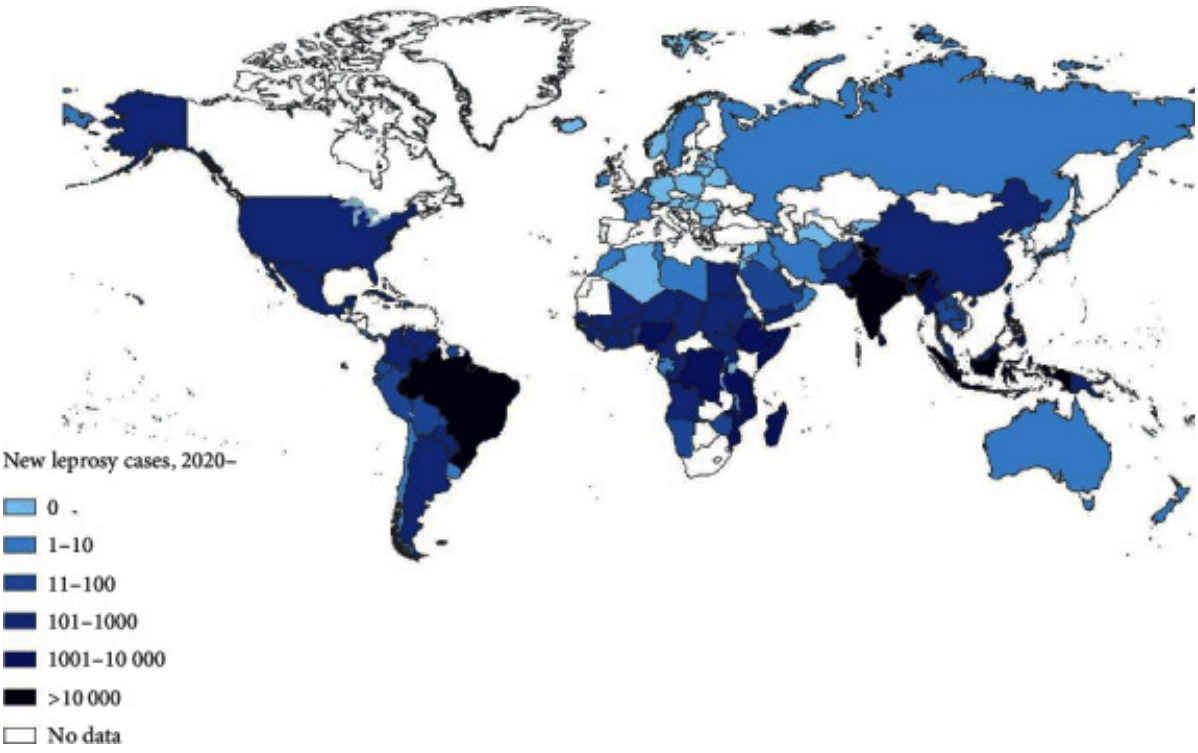


Figure 1.1: Global distribution of new leprosy cases in 2020 [2]



(a) Type I reactions
Top: erythema; **down:** skin lesion

(b) Type II reactions
Top: red nodules; **down:** skin lesion

(c) Lucio's phenomenon
Top/down: necrotic ulceration

Figure 1.2: Clinical leprosy symptoms of the different reaction types [13]

1.1.2. Diagnosis and treatment

Multidrug therapy (MDT) is proven to be an effective treatment when delivered before the infection consequences emerge [5]. Rifampicin, dapsone (diaminodiphenyl sulfone) and clofazimine are the three medications prescribed throughout a period of six months for PB patients and twice as long for MB patients. This drug combination aims to kill the pathogen and minimize the recurrence rate and occurrence of disease reactions. After starting the treatment, a leprosy patient is no longer contagious [1, 3].

Current leprosy diagnosing methods highly depend on the recognition of clinical manifestations, possibly combined with laboratory research for challenging cases [1]. According to WHO standards, the diagnosis relies on the identification of at least one of the key indicators: ”1) definite loss of sensation in a pale (hypopigmented) or reddish skin patch; 2) thickened or enlarged peripheral nerve, with loss of sensation and/or weakness of the muscles supplied by that nerve; 3) microscopic detection of bacilli in a slit-skin smear [3].” Early diagnosis of leprosy is still a challenge, and the impossibility of physical examination in early stages gives rise to a delay between disability onset and therapy [1].

Nerve function impairment is found as an indicator for the development of future immunological reactions or manifestations of motor and sensory neuropathy [15]. The voluntary muscle test (VMT) assesses the motor function of leprosy patients [16]. Sensory nerve function is typically evaluated using the monofilament test (MFT), which semi-quantitatively monitors touch sensation on the palms and soles [17, 18]. Brakel et al. (2008) [18] demonstrate a significant number of diagnosed leprosy patients either having or developing subclinical nerve damage not detected by the standard clinical tests VMT and MFT. According to Wilder-Smith et al. (1996) [19], autonomic nerve impairment (ANI) may serve as an identifiable measure of the risk for developing leprosy. Therefore, diagnosis based on pre-existing ANI seems to be promising for early-stage leprosy patients.

1.1.3. Infrared as a potential tool

Cavalheiro et al. (2016) [20] investigate thermography as a potential tool for diagnosing leprosy in an early stage and prove positive results. Infrared (IR) radiation, also called thermal radiation, arises because all objects with a temperature above 0 degrees emit electromagnetic radiation [21–23]. Peripheral neuropathy influences hand vasomotor control, resulting in a lowered temperature [20, 24]. Additionally, leprosy is correlated with deteriorated blood vessels in the limbs, which causes less blood flow in the distal extremities [6, 25]. These vascular transformations result in a reduced finger temperature and can be detected in leprosy patients before clinical symptoms are manifested [24, 26]. Medical infrared thermography (IRT) maps skin temperature non-invasively, enabling the demonstration of leprosy-related ANI in an early stage [27, 28]. Studies have also demonstrated valuable IRT diagnostic outcomes for other clinical conditions, such as neuropathy and diabetes [29–31]. Using this technique as a diagnostic tool will allow patients to start therapy earlier and will prevent deterioration of the disease and the associated symptoms [20].

1.1.4. Research initiative

The Nepalese healthcare institutions INF’s Green Pastures Hospital and Shining Hospital Banke have started a research initiative: Thermographic Assessment of Autonomic Impairment in Leprosy (TAIL). The project is in collaboration with the Leprosy Research Initiative (funding partner) and Delft University of Technology and led by the principal investigator S. Maharjan and co-investigators A. Knulst, W. Brandsma and C. Knulst-Verlaan. As stated in the research proposal: ”The general objective is to determine the value of infrared thermography (IRT) in the assessment of autonomic nerve function in newly diagnosed patients with leprosy. The clinical objective is to determine the functioning of the autonomic nerve system (ANS) and how it is related to sensory and motor nerve function (measured with VMT and sensory testing (ST)) in newly diagnosed patients with leprosy” [32].

Both leprosy patients and healthy subjects will participate in a cross-sectional, observational study. After applying a cold pressure test (CPT), an infrared video of the hands will be recorded to examine the peripheral autonomic impairment. The cold stimulus acts as an impulse to the autonomic nervous system and is used as a vasomotor reflex test [20, 33]. ”The cold pressure test has been used as a research method to test the integrity of the autonomic nervous system” Lamotte et al. (2021) [34]. A baseline measurement and recording of the skin temperature during 15 minutes after the CPT enable

proper temperature analysis for the TAIL project. Along with the extracted temperature data, the testing results of VMT, testing results of ST, demographics and clinical representation will be collected to achieve the research objective [32].

1.1.5. Manual temperature analysis

In order to execute the TAIL research, temperature data from the regions of interest (ROIs) that have been specifically chosen must be determined for further analysis. F. van den Bogaert developed a software program to manually annotate six ROIs per hand and extract the aimed temperature values from the IR videos. The customized Python model retains the one-time selected coordinates during the video and determines an average temperature within the hand regions. The results, including average temperature values and standard deviations, are further analysed in a separate MATLAB script. The outcomes of this analysis will ultimately be used to investigate the potential of infrared technology for leprosy diagnosis.

Manual ROI or landmark annotation has several disadvantages. Manual marking is time-consuming for the researcher or expert performing the task [35, 36]. According to Singh et al. (2017) [37], a manual approach has poor reproducibility, introducing non-standardisation and high variability in thermographic analysis. Therefore, an automatic method for ROI selection and temperature determination in the TAIL research is preferred.

1.2. Research objective

1.2.1. Automatic temperature analysis

This study aims to develop a method for automatic temperature analysis to explore the potential of infrared thermography for early-stage leprosy diagnosis. The custom-made model aims to support the TAIL research initiative conducted by the INF's Green Pastures Hospital and Shining Hospital Banke in Nepal. Diagnosing leprosy patients at an earlier stage with the help of IRT may hopefully reduce the harm caused by this neglected tropical disease in the future.

1.2.2. Model requirements

Based on the current knowledge and framework of the TAIL research, a set of criteria for the model has been established:

- The model should require as few as possible and at least fewer manual steps than the existing manual method.
- The model should be at least as reliable and accurate as the manual approach.
- No or as few additional expenses as possible should be incurred compared to the current process.
- The model should be easily implemented in the existing research environment.
- The model should be user-friendly for individuals without prior knowledge of the software.
- The model should not be bound by location restrictions.
- The model should be adaptable to accommodate alterations, such as when the input data or the intended output changes.
- The execution time of the model should remain within reasonable margins.

These requirements serve as guiding principles during the development process of the automatic approach for temperature analysis.

1.2.3. Study approach

The approach to achieving the goal of this study (1.2.1) involves adapting an existing machine learning model to ensure it is tailored to the specific context of application. The research objective includes improving the accuracy and efficiency of the temperature analysis of ROIs in the hands and validating the developed method. This involves the key steps of collecting IR data available from the TAIL research initiative and iterative training and testing of the model. The machine learning model is modified to enhance its performance for ROI detection. Subsequently, the temperature analysis within these localised regions is optimised. Additionally, validation is a critical part of this study. The same set of videos is

used to evaluate the model's performance and compare relevant outcomes to the results of the manual approach. Strategies to address challenges such as variations in video contrast and hand movement are considered in the development process. Finally, this thesis is structured as follows: Chapter 2 provides relevant background information, Chapter 3 details the methodology, Chapter 4 presents the results, Chapter 5 discusses the findings and future research recommendations, and Chapter 6 concludes the study.

2

Background

2.1. Related literature

The study ”Thermographic analysis and autonomic response in the hands of patients with leprosy” by Cavalheiro et al. (2016)[20] investigates the temperature changes in the hands of leprosy patients due to neuropathy using IRT and evaluates the autonomic response to cold stress. The observational, cross-sectional and controlled study design included n = 17 leprosy patients and n = 15 controls. Participants' hands were filmed with an infrared camera while resting their palms up for five minutes. A vasomotor reflex test was performed by applying a cold spray to the palms to observe the autonomic response. Temperature measurements were taken at specific points on the hands at the start, 2.5 minutes, and 5 minutes. The study found significant differences in temperature recovery between the control group and leprosy patients, using a significance level of 5%. The control group recovered 98.2% of the initial temperature after five minutes, while leprosy patients showed varied recovery rates depending on the clinical form of the disease. The results indicate impaired autonomic response in leprosy patients. Based on this study, IRT can effectively measure temperature changes in the hands of leprosy patients. This provides insights into the autonomic dysfunction associated with the disease. The findings highlight the potential of thermographic analysis as a diagnostic tool for assessing peripheral nerve function in leprosy patients [20]. The article does not provide a description of the performed annotation or segmentation of the specific points on the hands.

2.2. Manual temperature analysis

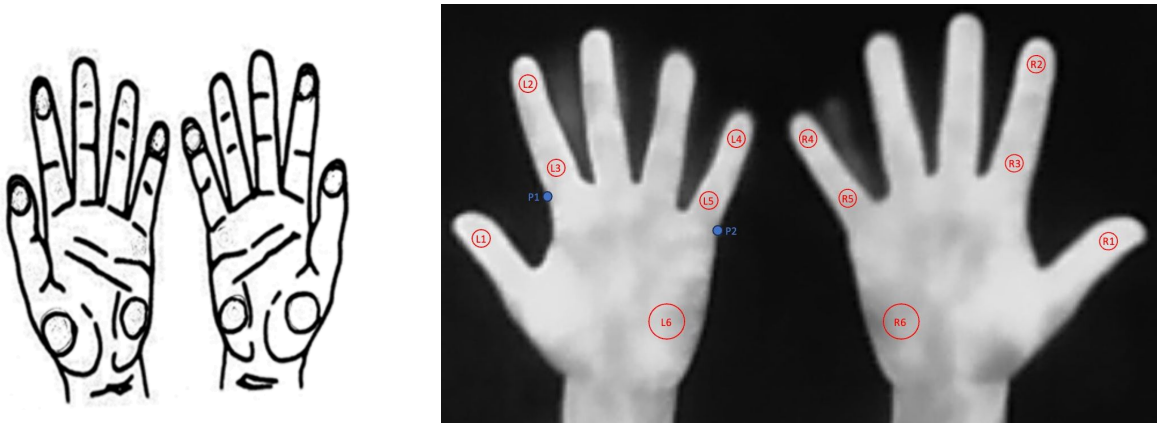
A manual approach to monitor the temperature progression of a selection of ROIs was constructed by F. van den Bogaert. The focus of this study will be analysing the same regions using a newly developed method. The results of the manual and automatic methods will be compared for validation.

2.2.1. Regions of interest

Based on the innervation areas of the median and ulnar nerve, a selection of six ROIs per hand is defined. Four similar regions, provided in Figure 2.1a, are used for sensory testing (ST) within the TAIL research (Section 1.1.4). Two additional areas located lower on the index finger and pinky had been included to assess the difference in distal and proximal finger temperature. The corresponding labels for the twelve ROIs are presented in the circular regions of Figure 2.1b. In the process of achieving the research objectives, the TAIL research initiative focuses solely on the index finger and pinky tips of both hands [32].

2.2.2. Algorithmic design

Figure 2.2 shows a schematic of the steps involved in the manual approach in Python. After loading the infrared video and creating an output folder, a window appears in which trackbars can be used to indicate the moments before cooling, directly after cooling, and the end of rewarming, i.e. after the hands have been recorded for fifteen minutes. The frame with the highest contrast is then used to man-



(a) Regions of interest for sensory testing **(b)** Twelve regions of interest with labels for the TAIL research

Figure 2.1: Visual demonstration of the regions of interest on both hands

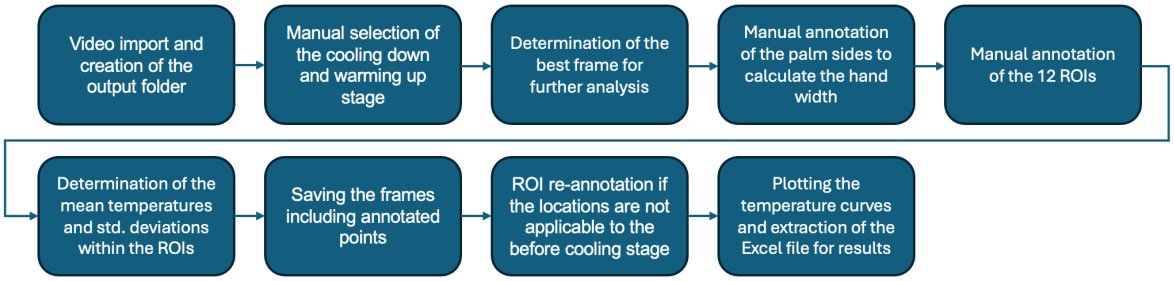


Figure 2.2: Schematic model overview: manual approach for temperature analysis

ually annotate the twelve ROIs in a graphical user interface. The size of the regions is determined by the distance between points P1 and P2, which are also manually selected. These points are illustrated on the left hand in Figure 2.1b. This size adjustment ensures that the method applies to varying hand sizes. For the regions located on the fingers (1 - 5), the ROI is a circle around the clickpoint with a radius of 10% of this distance. For the hand palm regions, L6 and R6, a circle with a 20% radius is employed. The temperature progression post-CPT is calculated every ten seconds for these twelve ROI locations, which are fixed throughout the video. The mean and standard deviation values within each region are saved in an Excel file, and plots depicting the values over time are extracted.

Since the ROI coordinates are annotated just once by the user of this method, hand fixation during the experiment is essential. Excessive hand movement causes the circles to be misplaced in other video frames, leading to incorrect temperature values being extracted. Additionally, if a circle partly extends outside the intended area, the divergent surrounding grey values can affect the results due to averaging. Besides the previously discussed disadvantages of a manual method versus an automatic one (Section 1.1.5), this compromises the reliability of this approach. It underscores the demand for real-time ROI annotation that can be automated. Sufficient real-time tracking could even eliminate the need for a fixation mould, potentially reducing material costs and experiment duration, increasing patient comfort, and allowing the experiment to be applied everywhere and for all types of hands, including those that are deformed.

3

Method

3.1. Data acquisition

At the INF's Green Pastures hospital, data is being acquired for the TAIL research initiative. The planned large-scale collection of infrared videos of the hands of both leprosy patients and healthy individuals enables the investigation of the potential of IRT for early diagnosis. An experimental setup and protocol have been established to generate data for the existing manual method. Along with newly produced videos using the same approach, this data was utilised for the development and validation of the automatic approach for temperature analysis. Whenever possible, subject details on demographics and clinical presentation are gathered. For this study, the development of an automatic model supporting the TAIL research initiative, these variables were not used.

3.1.1. Experimental setup

An experimental setup was designed to meet the following requirements:

- Ensuring consistent hand positioning both before and after the CPT, with the hands maintaining the same position for fifteen minutes post-test.
- Using a securely fixed camera throughout the recording process, preferably designed to withstand impact.
- Ensuring visibility of the entire skin surface of the (inside of the) hands, with no objects obstructing the camera's view.
- Providing a comfortable experience for patients during the fifteen-minute duration of the setup.
- Accommodating various hand sizes effectively.
- Ensuring the portability of the setup for use in different locations or hospitals.

The resulting setup and the four main steps from the experiment protocol are illustrated in Figure 3.1. While conducting the experiment, the hands are first placed in the mould to measure a baseline temperature. Secondly, the CPT is performed, after which the hands must be dried carefully to avoid reflection for the infrared camera. At last, the hands are filmed back in the mould for fifteen minutes to measure the response. The hand supports, custom-made from plastic moulds overlaid with foam, are secured to the setup using velcro. These supports facilitate personalised adjustments for each patient. The setup incorporates an infrared camera affixed for adjustable height, with connectivity to a phone via cable. Within these supports, hands and fingers are stabilised by inserting a foam block with velcro between the fingers, with a slightly larger block positioned between the thumb and index finger. This configuration enables precise repositioning of the hands in the holders after the CPT. Additionally, the setup is designed for easy portability, and patients have the option to hold it on their lap during the test for added comfort.



(a) The hands positioned in the mould for the baseline measurement



(b) The hands in cold water for one minute, to cool them down (CPT)



(c) Drying the hands after CPT to prepare for the measurement



(d) The hands back in the mould for response measurement

Figure 3.1: Experimental setup at the INF's Green Pastures Hospital

3.1.2. Cold pressure test

The CPT was performed using ice water. After the subject was acclimatised for fifteen minutes in the room, maintained at 24°C with air conditioning, the two hands were cooled simultaneously. The hands are completely immersed, up to the wrist, in two separate canisters containing 5°C water from the refrigerator. After one minute, the hands are sufficiently cooled. They can then be dried and placed back in the holder to continue the experiment.

3.1.3. Infrared camera

The infrared camera used for the experiments is the Uni-T UTi721M. The selection for this thermal camera was made for its affordable price, approximately €220, and due to its specifications that are adequate for the described application. In addition to the features mentioned in Table 3.1, the camera is compatible with an Android phone and has the ability to take videos [38]. White hot was the already chosen setting because it lent itself best to convert the pixel's grey values to temperature during video processing.

The recorded videos were stored in the data folder of the TAIL research initiative. Appendix A contains a comprehensive user guide for conducting the experiment and creating the video. For testing purposes, additional videos were recorded using variations of the previously described experiment setup.

3.2. Software selection

The software selection was based on specific criteria, as described in Section 1.2.2: Firstly, it was requested that no additional expenses were incurred in automating the temperature analysis. Additionally, it was required that the software be easily implementable within the current process. Python was chosen for these purposes, given its cost-free and open-source nature and the fact that it was already used in the previously developed manual method. The code was developed in Python 3.12.1 [39].

Table 3.1: Specifications of the Uni-T Thermal Camera [38]

Specification	Value
Brand	Uni-T
Hardware interface	USB Type-C
Supported file format	JPEG, MP4
Size	33 x 28.2 x 11.3 mm
Weight	12 g
IR resolution (pixels)	256 x 192
Spatial resolution	3.8 mrad
Field of view	56° x 42.2°
Temperature range	-20 to 550°C
Thermal sensitivity	50 mK @ 25°C
Frame rate	25 Hz
Accuracy	± 0.5° (30 ∼ 40°C)
Colour palettes	White hot, Lava, Ironbow, Black hot, Red hot, Rainbow, Rainbow HC

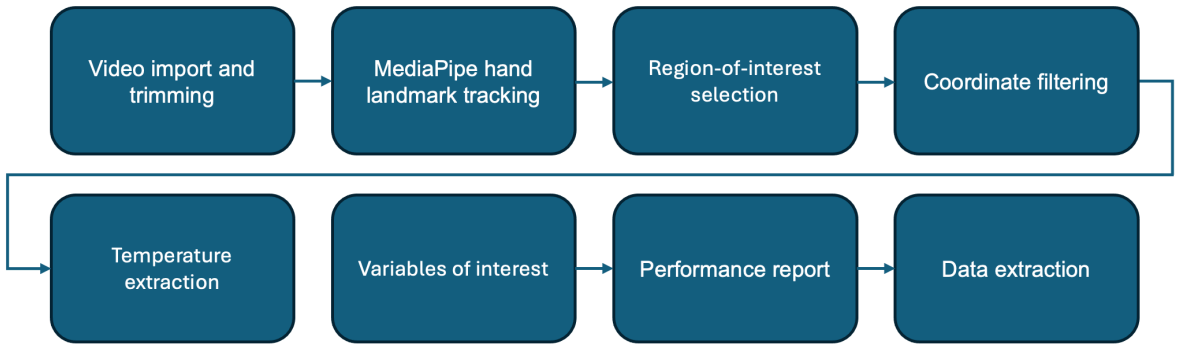


Figure 3.2: Schematic overview of the eight-step semi-automatic Python model for temperature analysis in infrared videos

3.3. Algorithmic design

The semi-automatic approach for temperature analysis consists of eight steps, schematically represented in Figure 3.2. These steps, starting with importing the data and ending with extracting the results, were all performed in Python. Subsections 3.3.1 to 3.3.8 delve deeper into the underlying principles and methodologies.

3.3.1. Video import and trimming

The initial part of the Python script handles the import of the required Python modules and the data, i.e. the infrared video to be analysed. Besides that, an output folder is created to store the outcome files. After activating the entire model, a File Explorer window automatically emerges to manually select the video. The output folder, named after the video, is generated in the correct directory. If the video already has an output folder, a new folder is created with a number starting from one appended to the video name.

The infrared video can be divided into three parts: the hands recorded before cooling, intended for the baseline measurement; the period during which the CPT is applied, with the hands out of view; and the hands after immersion in cold water, for approximately fifteen minutes during which the warming-up process is recorded. The first and third parts are of interest for further analysis, while the second is not and needs to be removed. Processing this segment adds extra computation time, and additionally, the measured temperatures of the background should not be wrongly included in the final data extraction. The infrared video is trimmed to create two separate videos, one of the before-cooling phase and one of the after-cooling phase. A window appears, containing a graphical user interface to select the start and endpoints of the two phases manually. Trackbars were created for user interaction, allowing the user to navigate through the video frames and select specific points for trimming (Figure 3.3). Subsequently,



Figure 3.3: Manual selection of the before-cooling and after-cooling phases using trackbars

four images of the keyframes, i.e. the start and end points of both experiment stages, are saved in the output folder for visual confirmation. It might occur that the last frame or a couple of frames of a video appear to be empty. Therefore, a function is created to find those empty frames and remove them in order to avoid errors during further processing. This is applied to both the initial video and the trimmed segments.

3.3.2. Mediapipe hand landmark tracking

The MediaPipe Hand Landmarker task, an existing and open-source machine learning model provided by Google, is the underlying principle for automatic ROI detection in this study. This Hand Landmarker solution from MediaPipe is able to detect 21 hand and finger landmarks in image or video data, as shown in Figure 3.4. The performed tracking is real-time and scalable to multiple hands simultaneously. Besides the image coordinates of these 21 hand knuckles, the left and right hand can be distinguished in case of more than one detected hand. The machine-learning model is trained on ±30.000 real-world images, in addition to various synthetic hand renders superimposed on different backgrounds [40, 41]. Mediapipe Hands solution is compatible with Python, among other platforms, and is available online at Google AI Edge: `https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker`.

The Hand Landmark system employs a machine-learning pipeline with two primary models: a palm detection model and a hand landmarks detection model. The palm detection model processes the

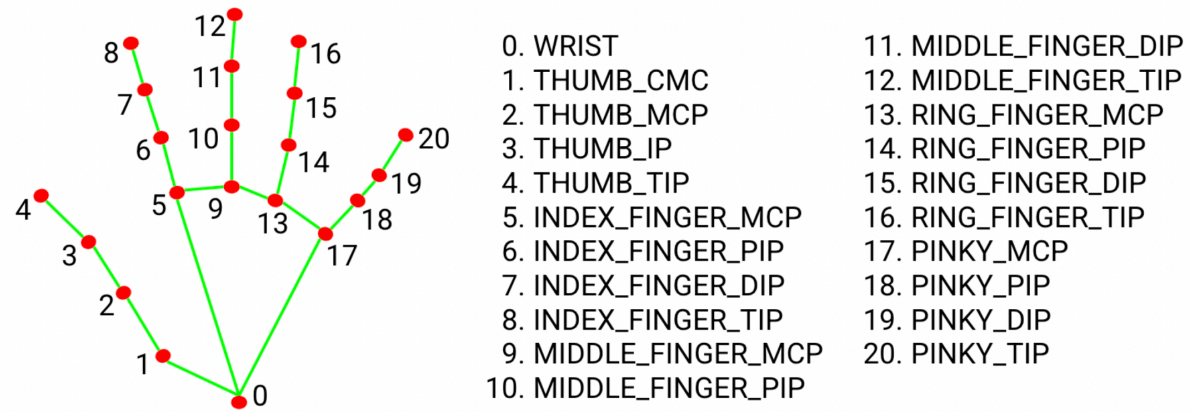


Figure 3.4: 21 landmarks from the MediaPipe hand landmark detection model [40]

entire image to locate hands, providing an oriented bounding box for the hands. The hand landmarks detection model then focuses on this cropped region to identify precise 3D hand key points. This approach minimises the need for extensive data augmentation, such as rotations, translations, and scaling, allowing the hand landmarks model to concentrate on accurately predicting coordinates. In video or live stream mode, the system enhances efficiency by using the bounding box from the hand landmarks model in one frame to predict hand regions in subsequent frames. The palm detection model is only reactivated if the hand landmarks model fails to detect or track the hands, thereby reducing the frequency of using the palm detection model and optimising performance [40, 41].

The downloaded *hand_landmarker.task* file is stored in the same folder as the Python code for the developed temperature analysis model. Importing the MediaPipe module in Python is required for the machine-learning task to be activated. Several configuration settings can be adjusted and implemented. The first option is the *static_image_mode*, which determines how input images are processed. When set to false, the system treats the input images as a continuous video stream, initially detecting hands in the first few images and then refining the localisation of hand landmarks. For subsequent images, once the maximum number of hands (*max_num_hands*) is detected and their landmarks are identified, it continues to track these landmarks without additional hand detection unless any hand is lost from tracking. This approach minimises latency and is suitable for video frame processing, as mentioned earlier [40, 41]. Conversely, when *static_image_mode* is set to true, the system performs hand detection on each input image independently, making it ideal for handling a batch of static, potentially unrelated images. The second option, *max_num_hands*, sets the upper limit on the number of hands to be detected in the input images. The third setting, *model_complexity*, defines the complexity level of the hand landmark model, which can be set to either 0 or 1. Higher complexity generally improves the accuracy of landmark detection but also increases inference latency. The min_detection_confidence parameter specifies the minimum confidence threshold (ranging from 0.0 to 1.0) required for the hand detection model to consider a detection successful. Finally, the *min_tracking_confidence* parameter indicates the minimum confidence threshold (ranging from 0.0 to 1.0) for the landmark-tracking model to consider the hand landmarks as successfully tracked. If the confidence falls below this threshold, the system will automatically re-invoke hand detection on the next input image. Setting a higher value for this parameter enhances the robustness of the solution but may increase discontinuation. This setting is ignored if *static_image_mode* is set to true, as hand detection is performed on every image in that mode [40, 41]. The values chosen for the configuration settings in this study are presented in Table 3.2.

The MediaPipe model outputs various details related to the detection and tracking of hands in given images or video streams. The primary output includes the multi_hand_landmarks, a collection of detected or tracked hands. Each hand within this collection is represented by a list of 21 specific hand landmarks, defined by three coordinates: x, y, and z. The x and y coordinates are normalized to the range [0.0, 1.0] relative to the image width and height, respectively. The z coordinate indicates the depth of the landmark relative to the wrist, which serves as the origin; a smaller z value suggests that the landmark is closer to the camera. Additionally, the model provides the multi_handedness, which

Table 3.2: Configuration settings of the implemented MediaPipe hand landmark task [40, 41]

Configuration setting	Possible value (range)	Implemented value
running_mode	True or False	False
max_num_hands	Any integer >0	2
model_complexity	0 or 1	1
min_detection_confidence	[0.0, 1.0]	0
min_tracking_confidence	[0.0, 1.0]	0.5

offers information on the handedness, i.e., left or right, of each detected hand. This feature comprises a label indicating whether the hand is left or right and a score representing the probability of the hand being labelled. The score is always 0.5 or higher, asserting a high confidence level in the handedness prediction. Lastly, the multi_hand_world_landmarks output details a list of 21 hand landmarks for each detected hand, but these are represented in world coordinates. This means each landmark has x, y, and z coordinates expressed in real-world 3D metrics (meters), with the origin set around the geometric centre of the hand [40, 41]. This latter output type and the z coordinates are not used in this study.

After applying real-time hand landmark detection to the before-cooling and after-cooling infrared videos, the landmark coordinates are stored in Python matrices. Subsequently, a set of n = 50 example frames with annotated landmarks is saved in the output folder. From the before cooling video n = 5 frames are saved: the first and last frame numbers, with three frames equally spaced in between. Similarly, for the after-cooling segment with a total of n = 45 images. The Mediapipe landmarks for the right hand are visible as blue dots, and those for the left hand as red dots.

3.3.3. Region of interest selection

For the purpose of the TAIL research initiative, it is not necessary to use all 42 MediaPipe landmarks, which include 21 for the left hand and 21 for the right hand. Reducing the number of coordinates and focusing only on the essential landmarks in the subsequent steps decreases computation time. Four new matrices are created to store the regions of interest coordinates: two for the x and y coordinates of both hands for the before-cooling phase and two for the x and y coordinates of both hands for after cooling. For labels L1 - L5 and R1 - R5 from the determined ROIs of the manual approach (Secion 2.2.1, Figure 2.1), the MediaPipe coordinates of thumb tip (4), index finger tip (8), index finger pip (6), pinky tip (20) and pinky pip (18) are used, respectively (Figure 3.3).

The hand palms, represented by labels L6 and R6 for the manual approach, cannot be directly obtained from the MediaPipe hand landmarks. Therefore, they need to be calculated using available information in the video frames. The palm coordinates are generated using the known coordinates of the wrist and the lower coordinate of the pinky, points 0 and 17 in Figure 3.4. The Euclidean distance between these two points is calculated. The x coordinate of the palm is positioned at two-thirds of this distance, measured from the wrist towards the pinky. For the y coordinate, one-third of the absolute difference between the y coordinates of these points is taken, again from the wrist towards the pinky. This places the palm coordinates for both hands, as illustrated for the left hand in Figure 3.5. The calculation is coded to work for hands in the regular orientation as well as when the hands are upside down in the recording.

The result of these steps are matrices for x or y values with ten rows for the extracted MediaPipe coordinates, alternating between the left and right hands, and two added rows for both palms. A list of the twelve corresponding labels is stored to ensure the correct labelling of calculated values and outcomes throughout the entire process. The columns of the four matrices represent the frames from either the before-cooling video or the after-cooling video.

3.3.4. Coordinate filtering

To gain insight into the locations of the twelve ROIs throughout the experiment, the x and y coordinates are plotted separately against time for the before and after cooling stages. Each line represents the movement of a specific landmark coordinate over time. The values range between 0 and 1, with the (0,0) coordinate located at the top-left corner of the video frames. The variable *fps*, i.e., frames per

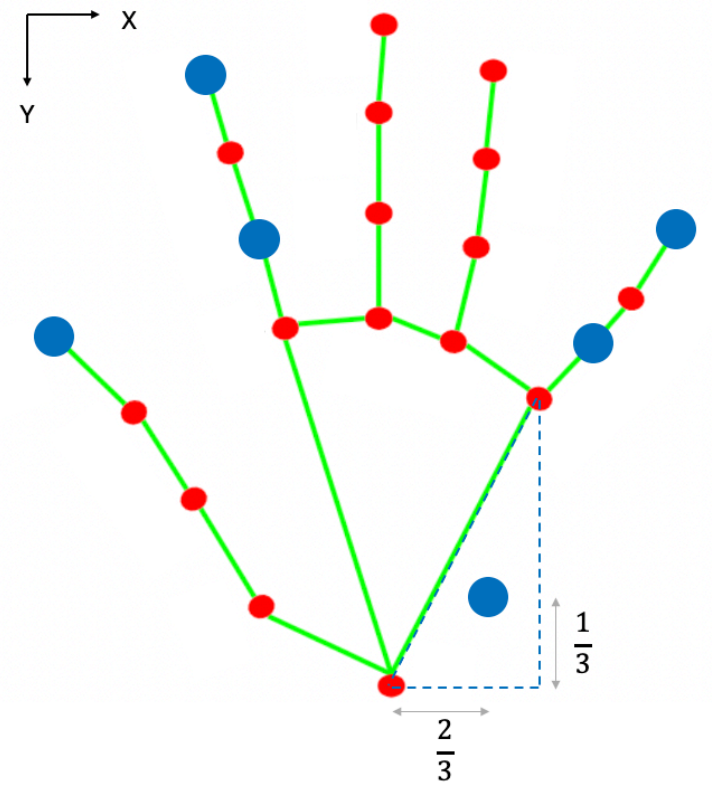


Figure 3.5: Schematic visualisation of the left-hand regions of interest selection

second, is used to convert the corresponding frame count into timesteps in seconds. These time values represent the progression of time throughout the video.

Due to the fixation used in the experimental setup, the hands remain relatively still. Consequently, the x and y coordinates should theoretically vary only within a certain margin during the procedure. However, because the tracking of the landmarks and, thus, the location of the ROIs is based on the MediaPipe model for hand tracking, this may not always be the case in practice. MediaPipe may not always detect a hand and its corresponding landmarks, resulting in matrices with NaN (not a number) values. Additionally, the resulting positions may also be inaccurately placed on the hands or even next to them. These two phenomena manifest as gaps or shifts in the coordinate plot lines. To address missing or incorrect coordinates, three sequential steps are implemented: interquartile range (IQR) filtering, interpolation of NaN values, and moving average smoothing. These steps were built upon the assumption that the hands remain relatively stationary during the experiment to accurately approach the ROI locations.

Interquartile range filtering

The first filtering technique utilises the IQR to exclude outliers, applied twice for enhanced robustness. The IQR is calculated as the difference between the lower quartile (Q1) and the upper quartile (Q3) of the data. Specifically, Q1 and Q3 represent the 25th and 75th percentiles, respectively, of the x or y coordinate values across the video. The upper and lower thresholds for filtering are determined as follows:

$$T_{min} = Q1 - c \times IQR \tag{3.1}$$

$$T_{max} = Q3 + c \times IQR \tag{3.2}$$

Here, T_{min} and T_{max} represent the lower and upper thresholds, respectively, and c is a factor typically set to 1.5 to ensure effective outlier removal [42, 43]. For the first application of the IQR filter, a factor c = 0 is used for the coordinate thresholds. Assuming that the majority of the coordinates found are

accurate, this method initially eliminates the most significant outliers and then subsequently addresses smaller remaining deviations [42–44].

Interpolation of NaN values
The four coordinate matrices, x and y values before and after, may contain NaNs for two reasons. Firstly, the MediaPipe model may fail to track hand landmarks in all frames of the videos. Secondly, outliers have been removed and replaced with NaNs through IQR filtering. Given that the fixation moulds suggest the coordinates do not vary significantly over time, the missing coordinates are not expected to differ much from the preceding or following values. Therefore, for each ROI, linear interpolation is applied to replace the NaNs with actual x and y coordinates and to resolve the gaps [45].

Before proceeding, the two steps mentioned above are repeated to eliminate the remaining outliers. In this second round, the rather standard value of c = 1.5 is applied to set the thresholds for the IQR filter.

Moving average smoothening
A moving average filter is applied to further smooth the data, reducing noise and providing a clearer representation of the trends in the landmark movements over time [46]. The average is calculated within a sliding window that moves across the coordinate data. This technique effectively smooths the data by averaging out the fluctuations and reducing the impact of irregularities. For each point, the moving average is computed by taking the mean of the values within the defined window centred around that point. The NaN values are excluded from the mean calculation to prevent them from affecting the average. If all values in the window are NaN, the smoothed data point is also set to NaN, indicating that no meaningful average could be computed for that position. Given dataset x, the moving average y at each point i is calculated using the following formula:

$$y[i] = \frac{1}{W} \sum_{j=-k}^{k} x[i+j] \tag{3.3}$$

Where:

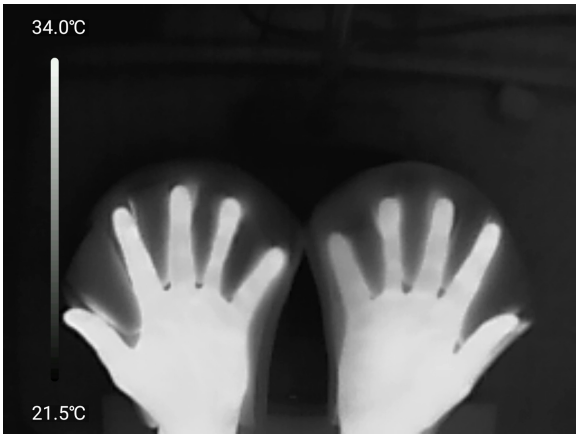
- W is the total number of observations within the window, which is $2k+1$.
- k is the number of observations on either side of i, making the window size extend from $i-k$ to $i+k$.
- $x[i+j]$ represents the data point j positions from i, with j ranging from $-k$ to $+k$.

After completing the three described steps, the resulting x and y coordinates are plotted against time. These lines show the effects of the filtering and smoothing techniques on the data, providing a cleaner visualisation of the landmark movements before and after cooling. Now that the final and filtered coordinates are known, the twelve ROIs can be displayed on example frames to provide visual insight. The regions of interest are defined by drawing a circle around the twelve stored (x,y) coordinates. To ensure that the circular ROI does not extend beyond the finger borders but still contains enough pixels for further temperature analysis, a radius of 10 pixels is set. A set of n = 50 frames from the video, again n = 5 for the before-cooling stage and n = 45 for after cooling, is saved in the output folder. The ROI boundaries for the left hand are indicated in red and for the right hand in blue. Besides, the frame number and time in seconds are indicated on each image.

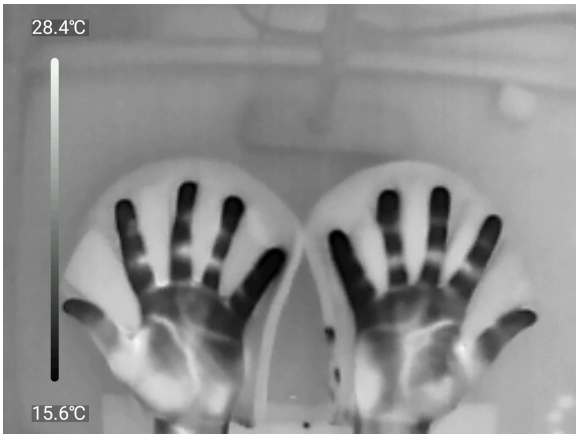
3.3.5. Temperature extraction

Since the locations of the ROIs in both hands have been determined, the temperature in these areas can be measured to study the temperature development post-CPT. The minimum and maximum temperatures linked to the colour scale, or rather grayscale, are not constant throughout the video (Figure 3.6). The scale ranges from black to white, representing cold to warm, respectively. However, for each frame, the specific minimum and maximum temperatures must be determined. This is achieved using Optical Character Recognition (OCR) implemented with the *Pytesseract* function in Python. To increase the likelihood of reading the correct values, a small region of pixels is specified in the frames where the function is applied. Due to poor contrast, it is not always possible to read a number correctly, or the determined value may be incorrect. To verify the latter, a check whether the minimum tempera-

ture remains below 30°C and the maximum temperature below 100°C is inserted. The scale minima and maxima are stored for each frame and may contain NaNs for unknown temperatures.



(a) Minimum 21.5°C, maximum: 34.0°C



(b) Minimum: 15.6°C, maximum: 28.4°C

Figure 3.6: Two frames from the same infrared video with different temperature scale values

Every tenth frame, approximately once per second, the temperature values of the twelve ROIs on the hands are determined. No critical information is lost by skipping the other frames, given the slow temperature changes during the warming up. Moreover, it reduces the running time of the model, as temperature determination by assessing every frame has a high computational load. The ROIs are considered circles with a radius of ten pixels around the centre coordinates. The pixel grey values within these areas are determined and converted into temperature values using linear transformation. The previously stored minimum and maximum temperatures serve as reference values for this transformation. To obtain the temperature values in a frame, the median is calculated for every ROI. This ensures that outliers, such as when an ROI approaches the edge of the finger, or temperature variations within the hand, have minimal influence on the computed temperature value. In case an ROI is partially crossing the frame boundaries or temperatures cannot be determined due to missing scale values, a NaN is assigned. Subsequently, the determined temperature development is plotted in two separate graphs against time. The first plot displays all ROIs collectively, while the second plot divides them into the left hand, right hand, and the difference between the same points on both hands. Based on the graphs, the user can get an initial impression of the hand rewarming after CPT. Additionally, any anomalies visualised in the plots may indicate insufficient reliability of the developed method. A new array for time intervals of ten frames is created to replace the original one-frame intervals.

3.3.6. Variables of interest

To explore the potential of infrared for early leprosy diagnosis, several variables of interest have been defined within the TAIL research initiative (Table 3.3). There are two types of existing variables: hand temperature values and recovery values. The first class represents the ROI temperatures before and at various time points after performing the CPT, along with the temperature drop due to cooling. The second class, temperature recovery, indicates the state of recovery as a percentage of the temperature drop after 5, 10, and 15 minutes. Additionally, the time measured to reach 50% and 95% recovery is of interest. The recovery variables may offer further insights into the impact of leprosy on the autonomic nervous system. Calculations based on the extracted ROI temperatures before cooling and after cooling were performed in Python to determine the variable outcomes.

3.3.7. Performance report

The developed semi-automated model for temperature analysis might have several imperfections. It is useful for the user to measure these in order to assess the method's reliability. The saved frames already aid in verification, allowing for visual inspection of the correctness of the ROI locations. Additionally, several other performance measures have been established to reflect the quality of the model.

Table 3.3: Extracted temperature and recovery variables

Variable	Description
Temp_Base	The temperature in an ROI before applying the CPT
Temp_CPT	The temperature in an ROI directly after the CPT
Temp_5	The temperature in an ROI at 5 minutes time after the CPT
Temp_10	The temperature in an ROI at 10 minutes time after the CPT
Temp_15	The temperature in an ROI at 15 minutes time after the CPT
Temp_Drop	The temperature drop in an ROI caused by the CPT, calculated from Temp_Base - Temp_CPT
Recovery_5	The recovery in an ROI at 5 minutes after the CPT, calculated as a percentage of Temp_Drop
Recovery_10	The recovery in an ROI at 10 minutes after the CPT, calculated as a percentage of Temp_Drop
Recovery_15	The recovery in an ROI at 15 minutes after the CPT, calculated as a percentage of Temp_Drop
t_50	The duration to reach 50% recovery of Temp_Drop in an ROI after the CPT
t_95	The duration to reach 95% recovery of Temp_Drop in an ROI after the CPT

1. Before cooling frames where left-hand coordinates could not be tracked [#, %]
2. Before cooling frames where right-hand coordinates could not be tracked [#, %]
3. After cooling frames where left-hand coordinates could not be tracked [#, %]
4. After cooling frames where right-hand coordinates could not be tracked [#, %]
5. Before cooling frames where the minimum temperature could not be read from the scale [#, %]
6. Before cooling frames where maximum temperature could not be read from the scale [#, %]
7. Before cooling frames where ROI (radius 10) exceeds the frame boundaries [#, %]
8. Before cooling frames where ROI temperature could not be determined [#, %]
9. After cooling frames where the minimum temperature could not be read from the scale [#, %]
10. After cooling frames where maximum temperature could not be read from the scale [#, %]
11. After cooling frames where ROI (radius 10) exceeds the frame boundaries [#, %]
12. After cooling frames where ROI temperature could not be determined [#, %]
13. Model running time [minutes]

The first four measures indicate how well MediaPipe was able to trace the hand landmarks, based on the absolute and relative number of frames with stored coordinates. This is examined separately for both hands in the before and after cooling video. It is important to realise that a MediaPipe placed landmark does not always indicate a correct location; therefore, it represents a quantitative rather than a qualitative performance metric. Items 4 through 8 relate to temperature extraction from the before-cooling part, and items 9 through 12 relate to the after-cooling part. These measures include the absolute and relative number of frames where the minimum or maximum temperature could not be read using optical character recognition, the number of frames where the temperature could not be determined due to frame boundary issues, and the total number of frames where the ROI temperature could not be determined. Ultimately, the total running time of the model is calculated and extracted.

3.3.8. Data export

In addition to the previously mentioned frames and plots that were stored, the quantitative results have been exported to an Excel file. A file is created in the output folder containing three tabs. The first tab includes the variables of interest for the twelve ROIs, together with an average for each variable (Section 3.3.6). The second tab accommodates the temperature values for the regions of interest, as calculated according to the description in Section 3.3.5. The initial row displays one temperature value for each ROI in the before-cooling phase. The median temperature is calculated over all assessed before-cooling frames, ensuring minimal impact from variance or outliers. Starting from the second row, the after cooling time steps and corresponding temperatures are listed. The final tab presents the model performance outcomes (Section 3.3.7).

3.4. Executive file

In order to enhance the usability of the developed Python code, it was transformed into an executable file. This modification allows users to operate the software without the need for direct interaction with the Python script or the Python environment. By converting the code into an executable format, users simply need to initiate the application, making it significantly more accessible for non-technical users. The conversion was achieved using *PyInstaller*. Appendix B provides a user guide on how to execute the semi-automatic temperature analysis on the infrared data.

3.5. Validation

Various analyses were conducted to validate the developed model for semi-automatic temperature analysis. These evaluations include the repeatability of the method, the accuracy of ROI localising based on visual examination and the effect of coordinate filtering. Additionally, the approach was compared with the manual method in a quantitative and qualitative manner. A few of the twelve ROIs, specifically the index finger tip and pinky tip on both hands, are considered the most important within the TAIL research. If necessary, the focus is limited to these four ROIs for the validation assessments. In addition to the videos where the experiment was recorded using the classical protocol, some videos were made with alternated test setups. These are used in a final examination to assess the quality of potential setup or method improvements for the future.

3.5.1. Output overview

To get an initial impression of the model's performance, the results from a set of infrared videos were compiled. The variables of interest, which were calculated after determining the temperature progression over time, are the outcome values that matter in the TAIL research. It is, therefore, crucial that these values be extracted using the developed method. To evaluate this, the number of known variables was calculated for each video. This is performed for each of the eleven variables from Table 3.3 and expressed as a percentage of the twelve ROIs. For example, when the baseline temperature can only be determined for the regions on one of the two hands, the outcome for this video will be 50% for Temp_Base. Subsequently, an average and standard deviation was calculated for every variable across the videos.

Besides the variables, the metrics from the performance report also provide information about the quality of the semi-automatic method. All results of the measures introduced in Section 3.3.7 have been organised, together with a calculated average and standard deviation. Possible explanations for the percentage of frames in which no ROI temperature was determined and for the total running time of the model (items 8, 12, 13) will be sought in the other metric outcomes, related to MediaPipe landmark tracking and temperature scale extraction. Furthermore, the results from the performance report will serve as a basis for seeking explanations within the various validation analyses.

3.5.2. Repeatability

One of the qualities measured in the validation analyses is model repeatability. In other words, how do the results compare when the model is tested multiple times on the same data. For a sample set consisting of half the classic recordings, two variants of this capacity were tested. The first is a test where the model is applied three times starting after the manual selection of the two experimental phases, before and after cooling. This examines the test-retest reliability of the algorithm's automatic part. Specifically, it assesses whether the implemented MediaPipe machine-learning task consistently outputs similar results when analysing the same frames multiple times. The second evaluation assessed the entire method twice, including the manual operation using trackbars. This introduces intra-observer variability, which can affect the final results of the semi-automated temperature analysis.

For the test-retest validation, the repeatability coefficient (RC) was used as a standardised measure to determine the absolute reliability [47]. The RC represents the threshold for repeated measurements taken under identical conditions. It determines the range within which the differences between these measurements are expected to fall with 95% probability. The RC can be calculated using the formula:

$$RC = 1.96 \times \sqrt{2}\sigma_w = 2.77\sigma_w \tag{3.4}$$

where σ_w is the within-subject standard deviation representing the standard deviation of values obtained from multiple measurements of the same variable [48]. After applying the method three times, the within-subject standard deviations are calculated for the extracted variables of interest. The evaluation to calculate the RCs is carried out in Python.

To assess whether intra-observer variability has a significant effect on the outcomes, the Intraclass Correlation Coefficient (ICC) is calculated. The ICC is a value that indicates the degree of reproducibility of measurements made by the same observer [49]. The value is determined using the following formula:

$$ICC = \frac{MS_B - MS_W}{MS_B + (k-1)MS_W} \tag{3.5}$$

Where:

- MS_B is the mean square between subjects, representing the variance of the average measurements between different infrared videos.
- MS_W is the mean square within subjects, i.e. the variation between two runs on the same file.
- k is the number of measurements [50].

The results are considered significant if the p-value <0,05, calculated using the F-test. "Based on the 95% confident interval of the ICC estimate, values less than 0.5, between 0.50 and 0.75, between 0.75 and 0.90, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively" Koo et al. (2016) [50]. The variables of the left index finger tip were used as input data. Similar to the test-retest evaluation, this reliability analysis is performed in Python.

3.5.3. Visual examination

In addition to all quantitative outcomes, it is valuable to qualitatively assess the accuracy of the determined ROI locations. Since the model itself is not capable of evaluating this, a visual assessment is necessary. Saved frames from the regular set-up videos were used to assign a visual examination score. Researcher I. Schemkes reviewed n = 5 before-cooling and n = 45 after-cooling frames that were stored in the output folder to determine whether each ROI was correctly annotated. The primary requirement was that the entire ROI should be within the respective finger or palm. Additionally, the circle for fingertip ROIs should fall within the upper half of the finger, and for finger pip ROIs within the lower half. After assigning scores of one for correct ROI placement and zero for incorrect placement, the totals were converted into percentages. This percentage indicates the number of correctly placed regions for each video and each ROI type. Average values and standard deviations are then calculated across both axes, i.e. per video file and for each of the twelve ROIs.

3.5.4. Effect of coordinate filtering

To provide a measure for the amount of coordinate filtering in each video, the Euclidean distances between the unfiltered and filtered ROI coordinates were calculated (see Formula 3.7). This distance represents the displacement in the image from the coordinates before and after the application of IQR filtering, NaN interpolation, and moving average smoothing. A mask was created to exclude NaNs in the unfiltered signal, for which no distance could be calculated. Subsequently, the average, maximum value and standard deviation of the Euclidean distances were computed [51].

$$Euclidean\ distance = \sqrt{(X_{unfiltered} - X_{filtered})^2 + (Y_{unfiltered} - Y_{filtered})^2} \tag{3.6}$$

Detailed analyses were conducted for the two infrared videos exhibiting the highest mean Euclidean distances. The initial addition involves comparing the ROI annotation accuracy between the MediaPipe landmarks and the ROIs after filtering. Given the duration of the after-cooling segment being approximately ten times longer compared to the before-cooling segment, the focus was placed on the after-cooling frames. Results from visual examination of the ROIs after filtering were already known, and for the MediaPipe landmark annotations, I. Schemkes applied the same visual examination technique to n = 45 stored post-CPT frames. For each ROI, the percentage of correctly placed regions is determined, along with an average and standard deviation across all ROIs. Since the palm regions

are added later in the process, this evaluation includes ten ROIs. Secondly, visual support consisting of three randomly selected moments during the experiment showcases frames with either hand landmarks or filtered ROIs.

3.5.5. Comparison with the manual approach

There is no ground truth available to validate the developed method for semi-automatic temperature analysis. However, the previously discussed manual approach, designed by F. van den Bogaert, is available as a reference. This method currently serves as standard within the TAIL research, but cannot be considered as absolute truth due to several drawbacks. In addition to comparing the approaches quantitatively with the variables of interest, it is, therefore, useful to provide more context by analysing the annotated frames. To compare the extracted temperature and recovery variables, they must first be determined for the manual method. The available infrared recordings have partly been analysed yet by a researcher from the TAIL initiative. The Excel files containing temperature profiles for the twelve ROIs, measured every ten seconds, were available for this study. A Python script has been written to calculate the eleven variables listed in Table 3.3.6 based on this temperature data. The emphasis is on the left and right index fingertips and pinky tips, as these regions are considered as most important within the TAIL research. To validate the results for these four ROIs and make a comparison between the manual method and semi-automatic method, Bland-Altman plots are used.

The Bland-Altman analysis is designed to examine the agreement between two different measurement methods. Specifically, "Bland and Altman stress the need to assess two aspects of agreement: how well the methods agree on average and how well the measurements agree for individuals" Bunce (2009) [52]. Per variable and for the four most important ROIs, the videos with known manual outcomes were examined. The mean differences were calculated, representing the average bias of the semi-automatic approach compared to the manual. A value close to zero suggests no significant bias measured. Subsequently, limits of agreement (LoA) were computed, based on the computed mean differences and corresponding standard deviations (SD):

$$LoA = Mean_{distance} \pm 1.96 \times SD_{distance} \tag{3.7}$$

These upper and lower limits provide a range within which 95% of the differences between the two methods' measurements are expected to lie. If the differences are consistently within these limits, the methods can be considered to have good agreement. However, if the LoAs are far apart, this indicates a large spread in the differences between the two methods and can suggest inconsistency in measurements. Eleven plots were created for all variables of interest, each displaying the calculated mean differences and LoA values as lines. The scatter points in different colours and symbols for specific ROIs and video files represent the differences for individual measurements against the average. Ideally, these points should be evenly distributed around the mean difference line with no pattern. If there is a trend where the differences increase or decrease with the magnitude of the measurements, it indicates proportional bias [52, 53].

These results were contextualised as well by adding a visual comparison. Researcher I. Schemkes selected two example videos based on the Bland-Altman plots: one video where the outcomes between the manual and semi-automatic methods visibly differ, indicated by points that deviate from the mean difference line or potentially approach or exceed one of LoAs, and a video where the scatter points were plotted relatively central. For each, three post-CPT frames were selected, and for comparison, the nearest corresponding frame available from the manual analysis is displayed.

3.5.6. Compare experiment set-ups

It is known that MediaPipe does not always successfully extract hand landmarks from infrared videos, which can result in NaN values within the matrices for x and y coordinates. To compensate for this, and for inaccurately annotated coordinates, filtering techniques were applied. Additional infrared videos were recorded with atypical experiment setups to test if the results from MediaPipe tracking, and consequently the outcomes of the model, could be improved. As described in Section 3.1.1, in the classical setup, the hands are held in position within a mould with foam attached. In some of the different experiments, the foam or even the entire plastic mould holding the hands is omitted. Besides improving

the outcomes of the semi-automatic method, this could also offer advantages such as saving time on installation and customization of the mould or enhancing patient comfort and convenience by allowing them to simply place their hands on the table. Lastly, there have been experiments performed using a mirror as the background beneath the hands, instead of the usual table surface.

To compare and validate different experimental setups, two main aspects were examined. Firstly, performance metrics related to MediaPipe's tracking were gathered from the performance reports. These include the proportion of frames lacking stored coordinates for both hands before and after cooling. Mean and standard deviation values were calculated for the performance measures across the three described setup variations. Additionally, a visual examination conducted by I. Schemkes focused on the n = 45 post-CPT frames — considered the most informative for evaluation — assessing the accuracy of ROI locations. Properly positioned ROIs suggest that the derived temperature and recovery variables of interest should likewise be accurate.

4

Results

4.1. Video details

A total of n = 27 infrared videos is included for this study, all of which meet the criteria that the complete experimental protocol was followed. Table 4.1 presents the characteristics of these mp4 files, namely the use of the fixation mould with or without foam, the type of background surface, and the availability of results from the manual analysis. For n = 15 videos, the results are known after manual analysis performed by one of the researchers of the TAIL research initiative. The pre-existing method developed by F. van den Bogaert has been utilised here. The experimental setup in these cases used a classical approach, with a fixation mould that included foam and a regular table as the background beneath the hands. These fifteen infrared videos were predominantly used for model development, training, generating results, and conducting validation analyses (Sections 4.2 - 4.3.4). The remaining videos are recordings following the regular experimental protocol but making use of atypical setups. In n = 5 of these cases, no fixation mould is used. Within this category, a mirror is used as the background surface in n = 2 instances. Additionally, a sample set of n = 7 recordings is created where the fixation mould is used without the foam attached to it. These twelve videos were used and compared to validate whether the various experimental setup designs led to better model performance compared to the classical setup (Section 4.3.5).

4.2. Model output

In the output folder created for each video, documents are saved as shown in Figure 4.1. The filenames starting with A contain saved frames that represent both experiment phases' start and end periods. Folder B holds frames with coloured MediaPipe hand landmarks, n = 5 from the before-cooling stage and n = 45 from the after-cooling stage. The x and y coordinate plots pre and post-filtering are stored in folder C, and the fifty frames with annotated ROIs in folder D. The filenames beginning with E contain four plots that show both unsmoothed and smoothed temperature developments over time. Additionally, the quantitative results are stored in the Excel file named "Results *videoname*", which includes three tabs for the variables of interest, temperature values over time, and the performance report. For the experiments utilising the classical setup, the primary outputs from the model are presented. Subsection 4.2.1 provides an overview of these n = 15 videos, and in Subsection 4.2.2 the key results from a selection of example videos are displayed. Given the substantial amount of results, a subset is presented in this section, with more extensive details available in Appendix C: In-depth Results.

4.2.1. Output overview

The quantitative outcomes of the n = 15 videos with a classic experimental setup have been compared. The resulting performance reports for this data are detailed in Table C.1, which can be found in Appendix C. For clarity in presentation, the metrics reviewed in the performance report are categorised into MediaPipe landmark tracking, temperature extraction, and model running time. Additionally, the variables of interest are included in Table C.2, also located in the same appendix.

Table 4.1: Characteristics of the infrared videos used for model training and validation

#	File name	Fixation mould [y/n]	Foam in mould [y/n]	Mirror background [y/n]	Manually analysed [y/n]
1	AN_11F_1	yes	no	no	no
2	AN_11F_2	yes	no	no	no
3	AP50F	yes	yes	no	yes
4	BG_44M_1	yes	no	no	no
5	BG_44M_2	yes	no	no	no
6	BG_44M_3	yes	no	no	no
7	BMT56F	yes	yes	no	yes
8	DCT_30F_1	yes	no	no	no
9	DCT_30F_2	yes	no	no	no
10	HD_SP23M	no	no	yes	no
11	HP_BM23M	no	no	yes	no
12	IKS14F	yes	yes	no	yes
13	KBS34M	yes	yes	no	yes
14	KP17M	yes	yes	no	yes
15	MS21F	yes	yes	no	yes
16	P_H_LIS26F	no	no	no	no
17	P_H_MAR24F	no	no	no	no
18	P_L_GAN45M	no	no	no	no
19	P-H-EB29M2	yes	yes	no	yes
20	P-H-R22M2	yes	yes	no	yes
21	P-H-RA29M2	yes	yes	no	yes
22	P-L-SB22M1	yes	yes	no	yes
23	P-L-SB22M2	yes	yes	no	yes
24	P-L-SUB22M3	yes	yes	no	yes
25	P-L-SUN29M3	yes	yes	no	yes
26	SBK22M	yes	yes	no	yes
27	SG30M	yes	yes	no	yes

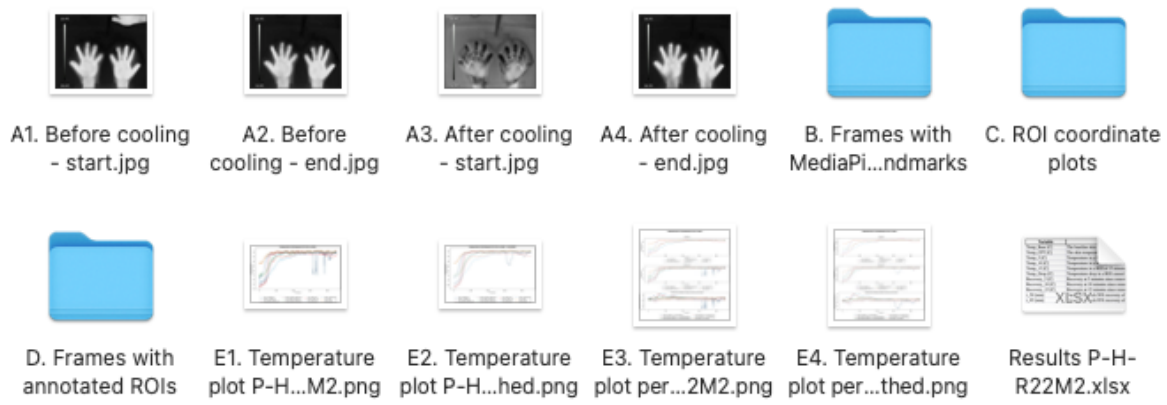


Figure 4.1: File structure of model output for example video P-H-R22M2

MediaPipe hand landmark tracking
In the performance reports, the MediaPipe tracking performance is calculated as the number of frames whose coordinates are not saved, expressed as a percentage of the total number of frames. For each video, the outcomes are presented in Table C.1. Across the n = 15 videos, the left-hand landmarks could not be tracked in an average of 14 ± 34% frames before cooling and in 17 ± 19% of the frames after cooling. For the right hand, these mean values are 15 ± 33% and 17 ± 21% before and after CPT, respectively. Notably, MediaPipe more frequently traces landmark coordinates in the complete before-cooling video segment compared to the after-cooling part. Both hands combined, this score of 0% frames with undefined landmarks occurs in n = 20 instances before and n = 3 after cooling.

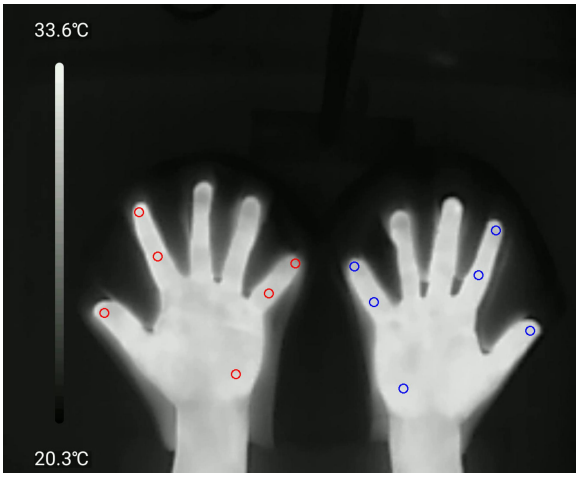
Temperature extraction
Four metrics in the performance reports represent the number of frames where either the minimum or maximum temperature could not be extracted using optical character recognition (Table C.1, labels 5, 6, 9, 10). The percentage of frames affected by this is below 5% in all cases. There is no negative effect from ROI coordinates being close to the edges of the frames for both experimental phases (labels 7, 11). According to column 8, for an average of 17 ± 35% of the before-cooling frames, the ROI temperature could not have been determined. In the cases of IKS14F and P-L-SB22M1, due to the complete absence of coordinates from one of the hands. For the after-cooling phase, this is 7 ± 7% (label 12). The percentages range from 2 to 26, with the absence of 100% outliers compared to the before-cooling frames. The highest value can be attributed to the same number of frames where the maximum temperature could not be extracted (label 10). In n = 4 of all cases, the percentage of frames where the minimum temperature could not be read can be linked to an identical percentage of undetermined ROI temperatures.

Model running time
The 13th column of Table 3.3.7 indicates an average model running time of 42 ± 15 minutes. As shown in Table C.1, after cooling videos with the highest number of untracked frames also exhibit longer running times. The maximum values include 78% frames for SBK22M, 62% for IKS14F, and 50% for P-L-SB22M1, with durations of 57, 61, and 73 minutes, respectively. In line with this, video P-H-R22M2, where landmarks were tracked in all frames, has the shortest running time of 26 minutes.

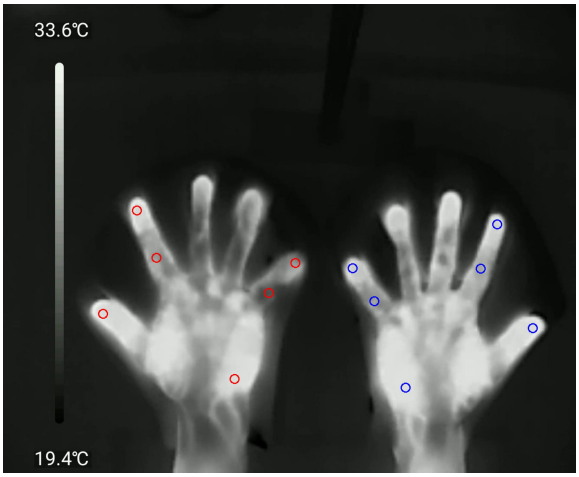
Variables of interest
For the temperature and recovery variables retrieved from the semi-automatic temperature analysis, a percentage is used to express for how many of the ROIs they could have been determined. In the vast majority of cases, the result is 100%, meaning that an outcome for the variable in question is known for twelve ROIs. However, in n = 2 cases, the baseline temperature, the difference between temperatures before and after CPT, and the recovery parameters could only be derived for 50% of the ROIs. For t_50 and t_95, which measure the time in minutes until 50% and 95% recovery is achieved, the average success rates are lower compared to the other variables, with averages of 87 ± 20 and 67 ± 29 minutes, respectively. In only n = 3 cases, the time to 95% recovery is known for each ROI. For the remaining n = 12 cases, the values vary between 25% and 92%, i.e., between three and eleven ROIs. 50% recovery (t_50) for all ROIs is achieved in n = 8 videos with the remaining percentages ranging from 33 to 92.

4.2.2. Example videos
The key results from a couple of videos are presented here to provide insight into the output of the developed semi-automatic model for temperature analysis. The first example is P-H-R22M2, file #20 from Table 4.1. Secondly, infrared recording AP50F, file #3 from the same table, is introduced. For more details and two extra examples, see the in-depth results in Appendix C.

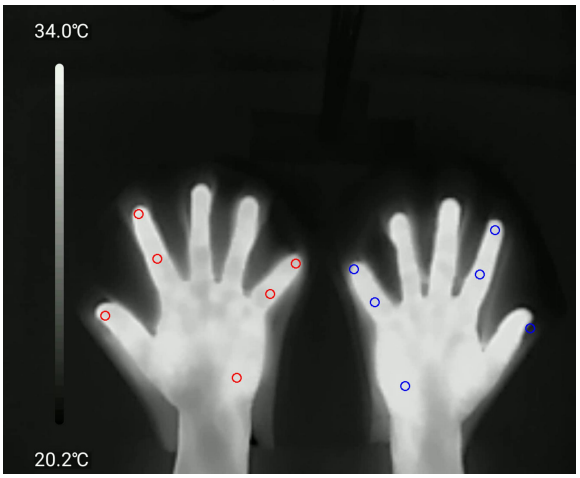
P-H-R22M2
Four frames showing the semi-automatically annotated ROIs of video P-H-R22M2 are presented in Figure 4.2. The temperature development plot can be found in Figure 4.3. The performance report, retrieved from Table C.1, indicates the following information: in all frames both before and after cooling, MediaPipe was able to trace the hand landmarks. For the before-cooling frames, all ROI temperature values could be determined. However, in 2% of the after-cooling frames, this was not the case. The minimum and maximum temperatures could not be extracted in 1% and 2% of the frames respectively. The model took 26 minutes to complete the analysis and the variables of interest outcomes, concerning temperature and recovery values, are given in Appendix C, Table C.3



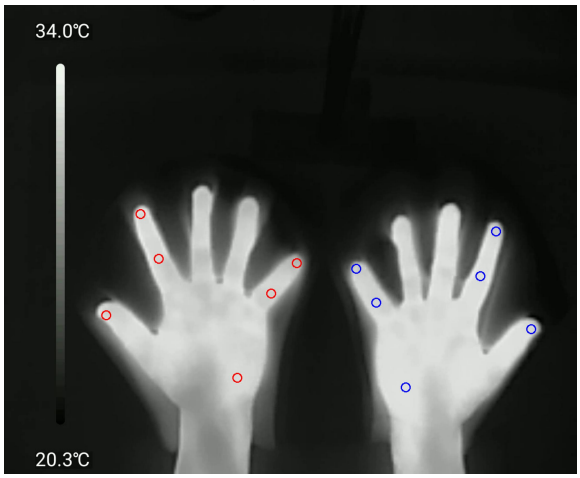
(a) Before cooling: frame 46, t = 4 sec



(b) After cooling: frame 1883, t = 144 sec



(c) After cooling: frame 8341, t = 639 sec



(d) After cooling: frame 11301, t = 866 sec

Figure 4.2: Annotated ROIs in frames of video P-H-R22M2: red for the left hand, blue for the right hand

AP50F
Similar example results are presented for video AP50F: Figures 4.4 and 4.5, along with Table C.4 in Appendix C, respectively display the annotated frames, temperature development, and variable outcomes. MediaPipe was unable to trace the left-hand coordinates in 93% of the before-cooling frames and in 53% of the after-cooling frames. For the right hand, these percentages were 4 pre-CPT and 12% post-CPT. Additionally, the minimum temperature could not be read in 1% of the after-cooling frames, and the maximum temperature in 5% of them. As a result, the ROI temperature could not be determined in 5% of the frames. Finally, the running time was 42 minutes (Table C.1).

4.3. Validation
In order to validate the outcomes of the model (Section 4.2) and assess the quality of the semi-automatic temperature analysis, several evaluation assessments were conducted. For the n = 15 IR videos, recorded using the standard experimental setup, analyses included tests for repeatability, visual examinations, assessments of coordinate filtering effects, and comparisons with the manual method (Sections 4.3.1-4.3.4). For the final validation, which involved comparing different experimental setups, the n = 12 alternative experiment files were used (Section 4.3.5).

4.3.1. Repeatability
In a sample of n = 7 videos, recorded under normal circumstances, the model was tested three times under identical conditions for each video. Due to identical variables of interest outcomes across all three

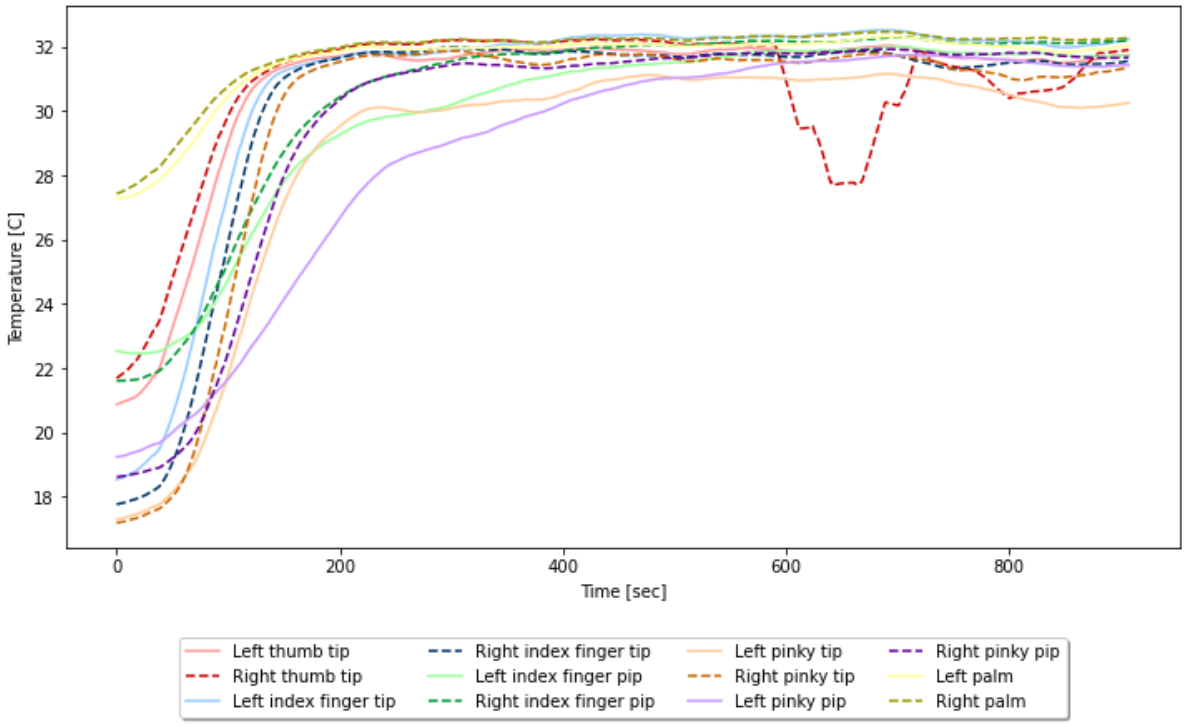
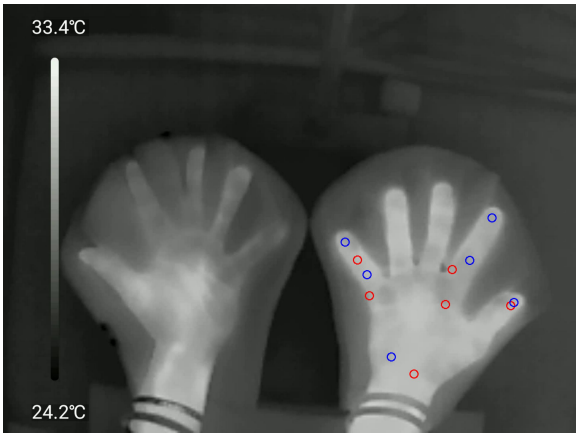
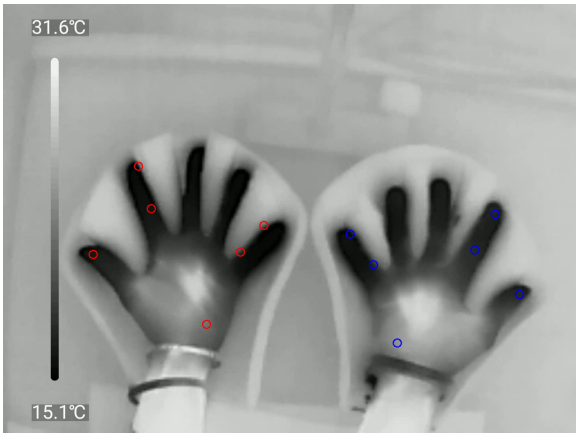


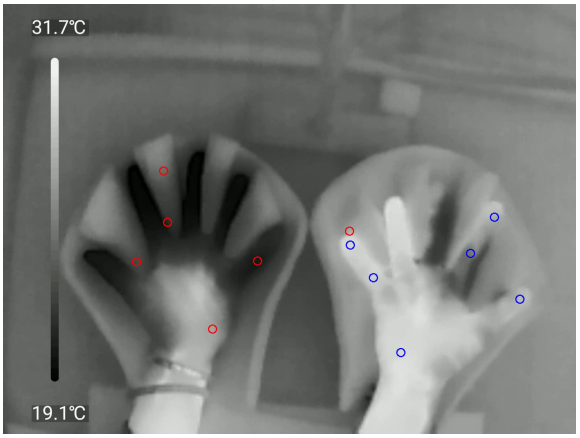
Figure 4.3: Temperature development over time for the twelve ROIs of video P-H-R22M2



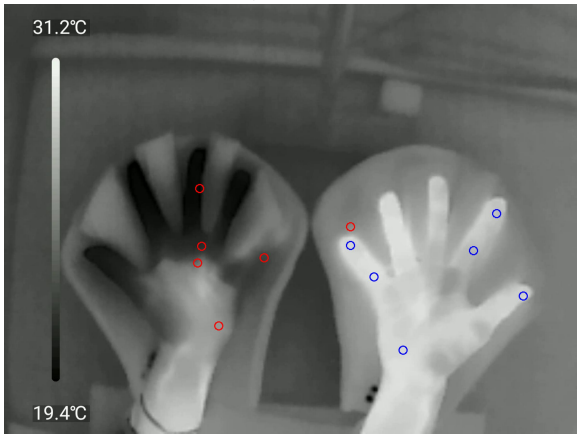
(a) Before cooling: frame 84, t = 7 sec



(b) After cooling: frame 1561, t = 131 sec



(c) After cooling: frame 7805, t = 656 sec



(d) After cooling: frame 11187, t = 940 sec

Figure 4.4: Annotated ROIs in frames of video AP50F: red for the left hand, blue for the right hand

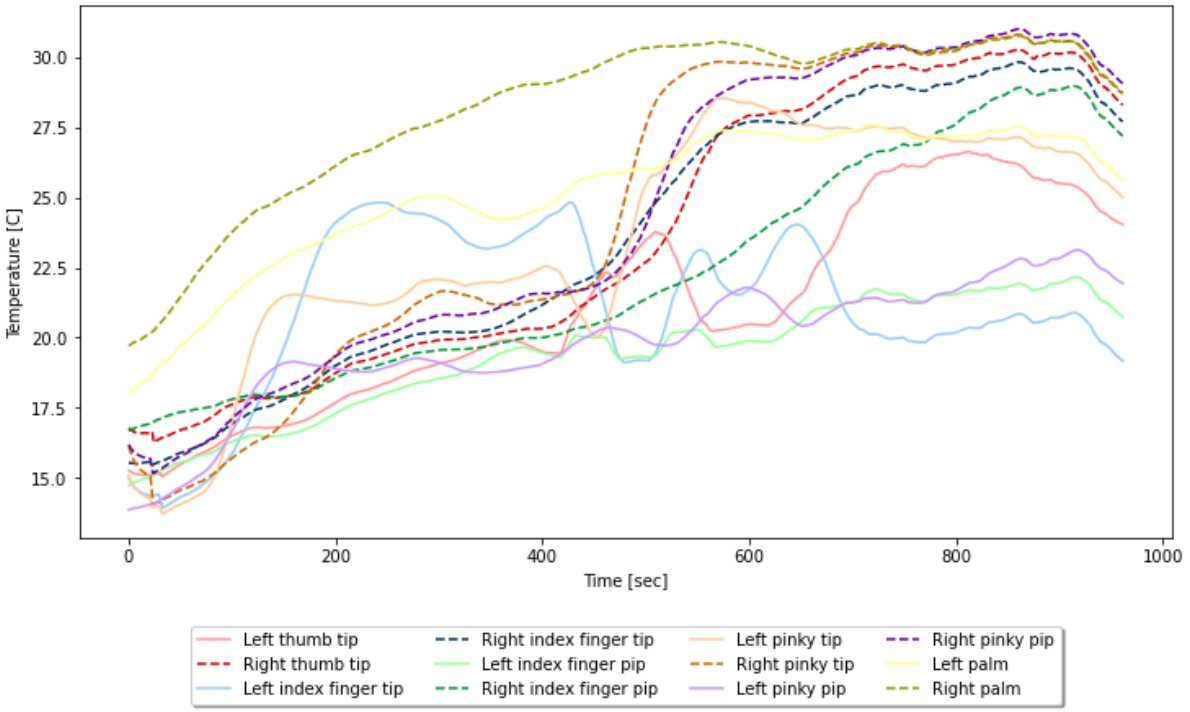


Figure 4.5: Temperature development over time for the twelve ROIs of video AP50F

runs for each video, the within-subject standard deviations (σ_w) were found to be zero. Consequently, the repeatability coefficient, calculated as $2.77\sigma_w$, was also zero in every case.

For the eleven variables of interest, ICCs and p-values were determined to assess the intra-observer variability within the same subset of seven infrared videos. The results are given in Table C.5 (Appendix C). On average, the ICC was 0.997 ± 0.002 with a p-value < 0.001 for all variables.

4.3.2. Visual examination

Visual examination is performed on the saved frames from both experiment stages: n = 5 for the before-cooling stage and n = 45 frames for the after-cooling stage. The before-cooling results can be found in Table C.6, Appendix C, and the after-cooling outcomes in Table 4.2. For the set of n = 15 regular recordings, as described in section 4.1, the percentage of correctly placed ROIs is determined. For example, for file AP50F, 56%, i.e., 25 of the after-cooling frames have the left thumb tip correctly annotated by the model (Table 4.2).

For the before-cooling phases, all ROI coordinates are correctly identified in n = 9 videos. In n = 3 instances, either the left or right hand is completely missing, resulting in an outcome of 0%. In one of these case, video AP50F, a percentage of 93 untracked landmarks coordinates was observed (Table C.1). Based on the bad results of the visual examination, it can be concluded that the attempt to reconstruct the coordinates has not been successful. As seen in Figure 4.4a, the red-coloured circular ROIs are not correctly localised. In addition to the zero scores, there are some cases where several ROIs of the left hand are partially or incorrectly placed. In summary, this results in an average of 87 ± 7% of ROIs being correctly annotated pre-CPT, according to the visual examination performed by I. Schemkes.

After cooling, the percentage of correctly annotated frames for the twelve ROIs averages 75 ± 8%. The mean and standard deviations are provided for both the videos, presented in the table rows, and the ROIs, in the columns. In n = 4 files, a mean percentage of 100 is achieved. The lowest success rates are for the videos MS21F, SBK22M and IKS14F, with averages of 0 ± 0%, 17 ± 34% and 37± 47%, respectively. This value range indicate the widest possible disparity. According to the performance report of MS21F, MediaPipe annotations are relatively similar to those of video P-H-RA29M2 with a maximum visual score. This example does not indicate a direct cause-effect relationship between the

Table 4.2: Percentage of after-cooling frames with correctly annotated ROIs, based on visual examination

	Left thumb tip	Left index finger tip	Left index finger pip	Left pinky tip	Left pinky pip	Left palm	Right thumb tip	Right index finger tip	Right index finger pip	Right pinky tip	Right pinky pip	Right palm	Mean ± Std. Deviation
AP50F	56	18	44	11	80	100	100	100	100	71	71	100	71 ± 33
BMT56F	100	100	100	100	100	100	96	100	100	100	100	100	100 ± 1
IKS14F	100	98	100	20	31	100	0	0	0	0	0	0	37 ± 47
KBS34M	100	100	100	100	100	100	100	100	100	0	7	100	84 ± 38
KP17M	100	100	100	67	73	100	100	100	100	100	100	100	95 ± 12
MS21F	0	0	0	0	0	0	0	0	0	0	0	0	0 ± 0
P-H-EB29M2	100	64	100	100	100	100	100	100	100	100	100	100	97 ± 10
P-H-R22M2	100	100	100	100	100	100	96	100	100	100	100	100	100 ± 1
P-H-RA29M2	100	98	100	100	100	100	100	100	100	100	100	100	100 ± 1
P-L-SB22M1	100	100	100	100	100	100	38	38	38	38	38	38	69 ± 32
P-L-SB22M2	100	100	100	98	100	100	100	100	100	100	100	100	100 ± 1
P-L-SUB22M3	96	100	100	0	0	100	76	36	36	96	100	100	70 ± 40
P-L-SUN29M3	98	100	100	100	100	100	47	91	98	87	98	98	93 ± 15
SBK22M	0	0	0	0	0	0	78	4	16	0	9	100	17 ± 34
SG30M	82	100	100	69	53	100	56	100	100	98	100	100	88 ± 18
Mean ± Std. Deviation	82 ± 35	79 ± 39	83 ± 37	64 ± 44	69 ± 41	87 ± 35	72 ± 36	71 ± 42	72 ± 41	66 ± 44	68 ± 44	82 ± 37	75 ± 8

performance of landmark tracking and the success of visually observed ROI locations. Conversely, the two other low accuracy scores from files SBK22M and IKS14F do contain the highest percentages of untracked after-cooling frames, namely 62 en 78 for left and right, respectively. Across the fifteen videos, the twelve ROIs have mean percentages ranging between 64 ± 44% (left pinky tip) and 87 ± 35% (left palm). The palms on both hands exhibit the highest percentages among the six ROIs. The pinky tips score the lowest, with mean values of 64% and 66% for the left and right hands, respectively.

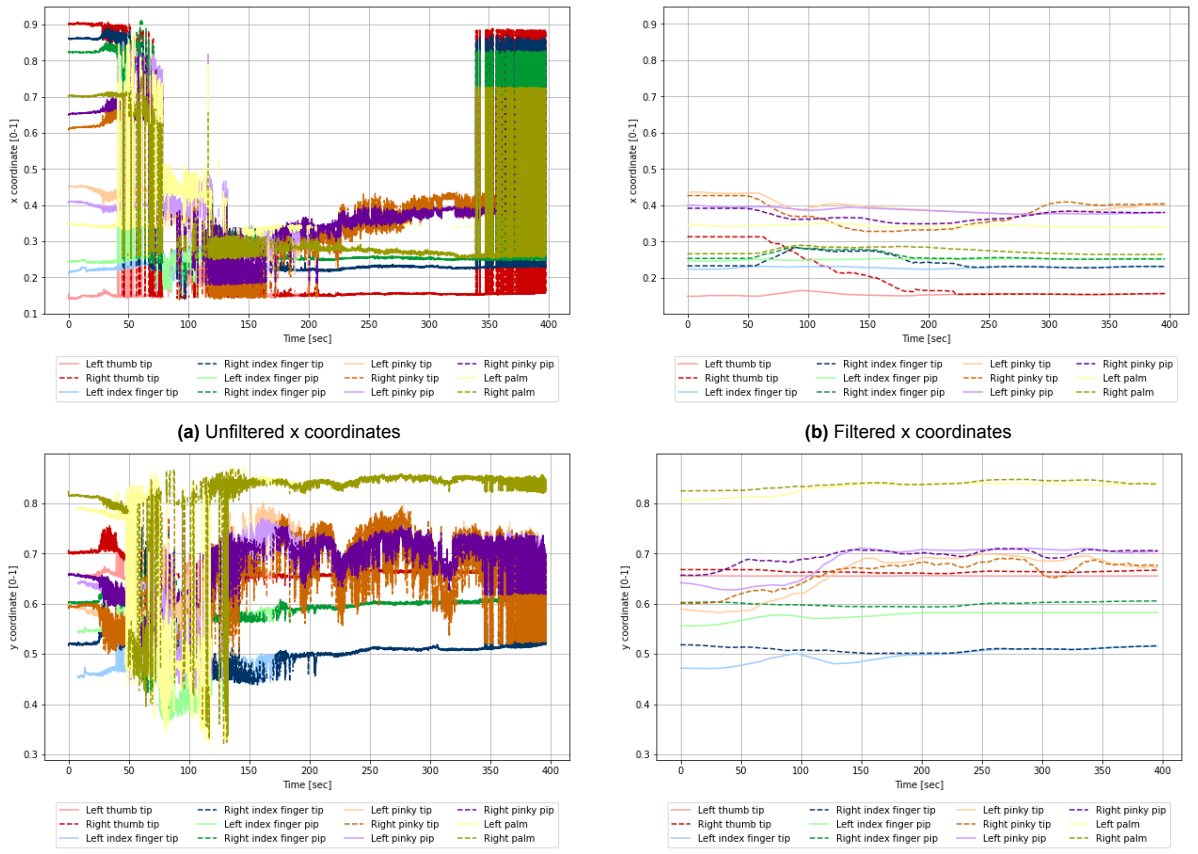
4.3.3. Effect of coordinate filtering

For the videos derived from the classical experiment setup, the Euclidean distances between the unfiltered and filtered ROI coordinates were calculated. These values serve as a measure for displacement after IQR filtering, linear interpolation, and moving average smoothing. Detailed statistics such as the mean, maximum distances, and their standard deviations are summarised in Table C.7, located in Appendix C. Among the n = 15 videos analysed, the average distance measured between the MediaPipe landmarks and the post-filter ROI coordinates is 0.03 ± 0.03. The maximum Euclidean distances measured vary from 0.39 to 0.80. The videos IKS14F and P-H-RA29M2, which displayed the largest mean distances of 0.11 and 0.07 respectively, were selected for further visual assessment. A remarkable fact is that within this subset, these videos have the highest and lowest average scores for visual assessment, i.e. 37% and 100%, respectively (Table 4.2). In terms of MediaPipe tracking performance, Table C.1 shows that the first-mentioned video has 62% untracked frames for the left hand, approximately three times higher than the average of 17% per hand. For the right hand, the difference to the average is smaller, with a measured value of 20%. For P-H-RA29M2, the left hand is well below the mean value with only 4% untracked frames, whereas for the right hand, landmarks cannot be found in 25% of the frames by MediaPipe.

IKS14F

In sample video IKS14F, the average Euclidean distance between the MediaPipe landmarks and ROIs after filtering is 0.11, with a standard deviation of 0.18 and a maximum distance of 0.75. Figure 4.6 presents the plotting of the x and y coordinates over time for the after-cooling stage, both unfiltered and filtered. The results from visual examination, i.e. the percentage of correctly annotated ROIs, are provided in Table 4.3. On average, the MediaPipe landmarks are correctly annotated in 16 ± 10% of the frames. After application of the IQR filters, linear interpolation and moving average smoothing, this value increases to an average of 35 ± 46%. At three timesamples post-CPT, frames showing either annotated MediaPipe landmarks or ROI regions are displayed as examples (Figure 4.7).

The plots of x and y coordinates before filtering, shown in Figures 4.6a and d, display considerable fluctuations for video IKS14F. The filtered lines appear much more stable with reduced variation over



(a) Unfiltered x coordinates

(b) Filtered x coordinates

(c) Unfiltered y coordinates

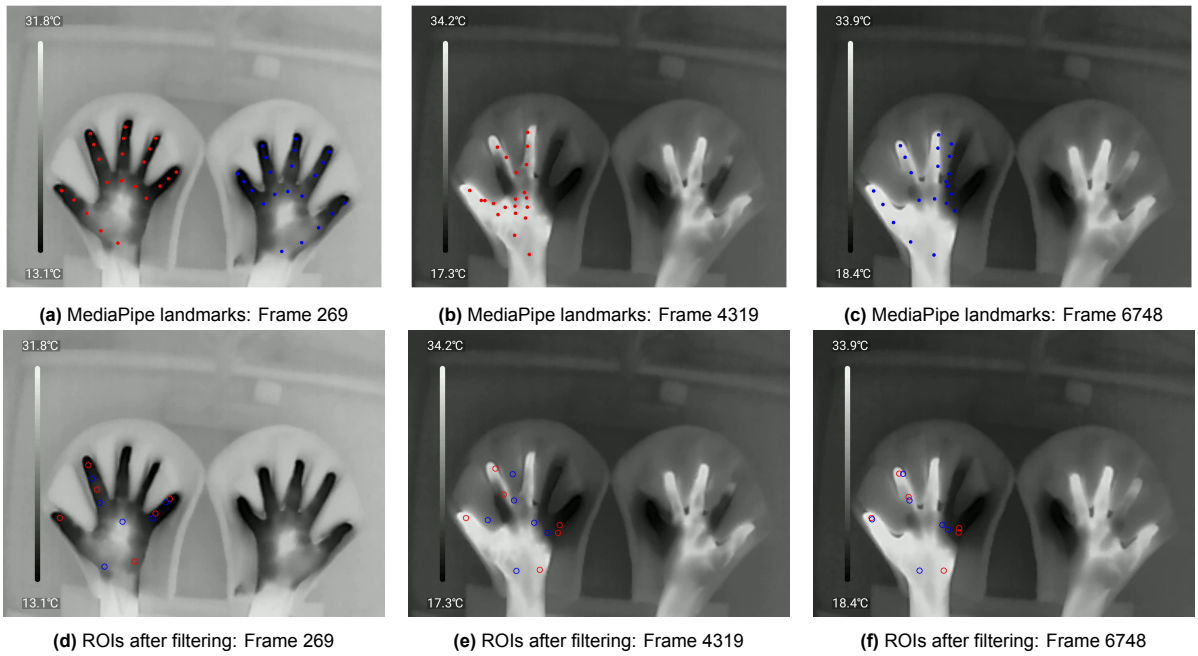
(d) Filtered y coordinates

Figure 4.6: After-cooling unfiltered and filtered ROI coordinate plots for video IKS14F

Table 4.3: Percentage of after-cooling frames with correctly annotated MediaPipe landmarks or ROIs, based on visual examination of video IKS14F

	Left thumb tip	Left index finger tip	Left index finger pip	Left pinky tip	Left pinky pip	Right thumb tip	Right index finger tip	Right index finger pip	Right pinky tip	Right pinky pip	Mean ± Std. Deviation
MediaPipe landmarks	31	27	27	9	13	11	16	16	2	7	16 ± 10
ROIs after filtering	100	98	100	20	31	0	0	0	0	0	37 ± 47

time. For the x coordinates, the range has significantly narrowed, and the lines for the left and right hands intersect, suggesting an overlap in the image, which is not feasible due to the use of a fixation mould. The results from the visual examination show an increase in correctly annotated frames from 16% to 37% after filtering (Table 4.7). For all left-hand ROIs, the accuracy score increases with differences ranging between 11% and 73%. However, for the right hand, the percentages, which initially lay between 2 and 16, all drop to zero. This indicates that these five regions are not correctly positioned in any assessed frame. The examples in the second row of Figure 4.7 confirm these findings: in a part of the video frames the pinky ROIs of the left hand have failed, and the right-hand regions are all located near the left hand. In frame 269, the unfiltered coordinates of the MediaPipe landmarks are still correctly positioned (Figure 4.7a). The disruptions that follow at a later stage, clearly visible in the unfiltered coordinate plots and the MediaPipe frames in Figures 4.7b and 4.7c, dominate and result in poorer outcomes of the right-hand ROIs after coordinate filtering. For the left hand, the landmark locations tracked by MediaPipe's machine-learning task sufficiently match reality to improve the result with IQR filtering, NaN interpolation, and moving average smoothing. Notably, the percentage of frames in which MediaPipe failed to localise hand landmarks was approximately three times higher for the left hand (62%) compared to the right hand (20%), according to the performance report (Table C.1).



(a) MediaPipe landmarks: Frame 269 **(b)** MediaPipe landmarks: Frame 4319 **(c)** MediaPipe landmarks: Frame 6748

(d) ROIs after filtering: Frame 269 **(e)** ROIs after filtering: Frame 4319 **(f)** ROIs after filtering: Frame 6748

Figure 4.7: Visual comparison between the MediaPipe landmarks and the final ROI locations at three moments in video IKS14F

Table 4.4: Percentage of after-cooling frames with correctly annotated MediaPipe landmarks or ROIs, based on visual examination of video P-H-RA29M2

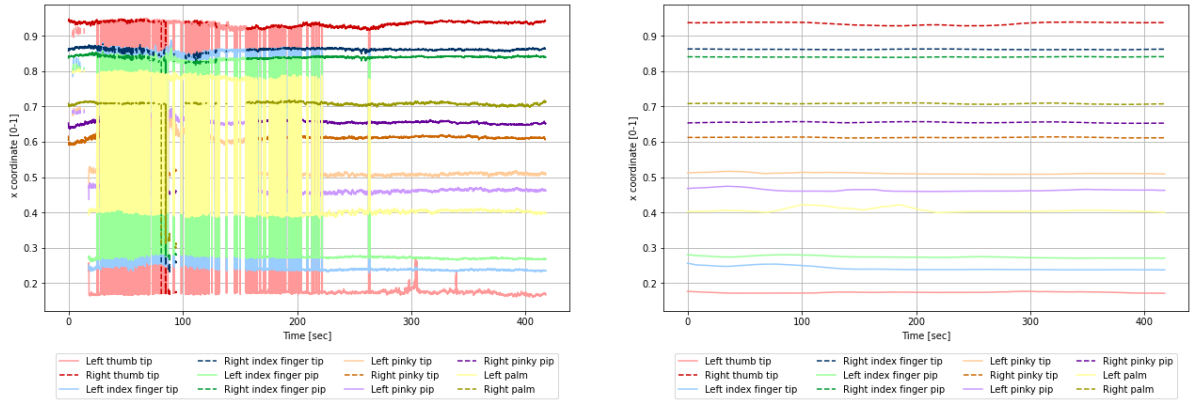
	Left thumb tip	Left index finger tip	Left index finger pip	Left pinky tip	Left pinky pip	Right thumb tip	Right index finger tip	Right index finger pip	Right pinky tip	Right pinky pip	Mean ± Std. Deviation
MediaPipe landmarks	62	64	64	64	64	71	67	67	69	69	66 ± 3
ROIs after filtering	100	98	100	100	100	100	100	100	100	100	100 ± 1

P-H-RA29M2

File P-H-RA29M2 exhibits the second-highest average Euclidean distance, recorded at 0.07. The associated standard deviation is 0.18, and the maximum distance observed is 0.78. The coordinate plots over time can be found in Figure 4.8. Compared to the previous example, the lines before filtering appear less cluttered, and despite the presence of outliers, a clearer trend can be discerned for each ROI. Again, the filtered signals are more stable, with a clear separation between the left hand [~0.25-0.55] and the right hand [~0.6-0.95] in subplot 4.8b, and the same ROIs for both hands approximately at the same level in subplot 4.8d. Visual examination indicates an improvement from an average of 66 ± 3% of frames to 100 ± 1% of frames with correctly located ROIs (Table 4.4). This is confirmed in Figure 4.9: where in an example frame 284, the left hand is missing, and in 1138 neither the right hand is measured nor are the left-hand landmarks on the right hand, all ROIs are correctly annotated after filtering.

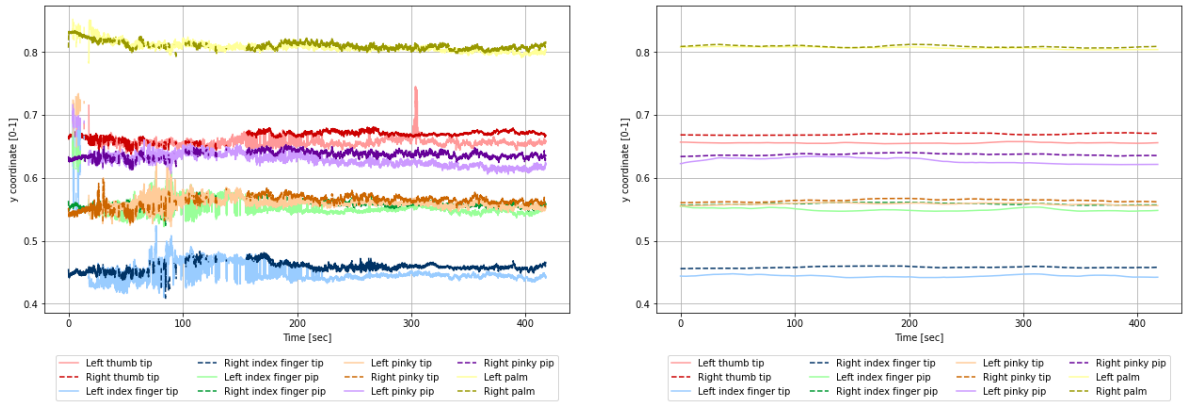
4.3.4. Comparison to the manual method

The outcomes of the semi-automatic method were compared with those of the pre-existing and current standard: the manual approach designed by F. van den Bogaert. For the eleven variables, n = 15 videos were analysed with a focus on the four most relevant ROIs. The Bland-Altman plots, which graph the differences between the manual and automatic methods against the average of these two methods for each measurement, are presented in Figure 4.5. The dashed black lines represent the mean difference, while the grey lines indicate the limits of agreement (LoAs). These limits provide a range within which 95% of the differences between the two methods are expected to lie.



(a) Unfiltered x coordinates

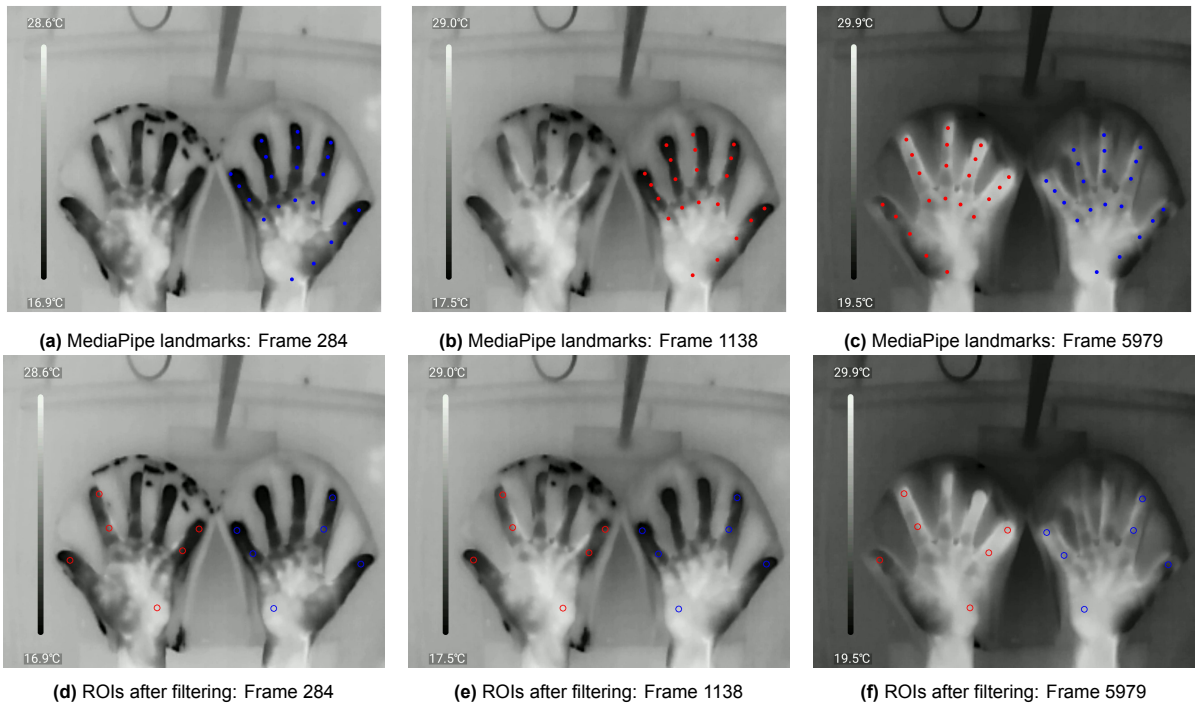
(b) Filtered x coordinates



(c) Unfiltered y coordinates

(d) Filtered y coordinates

Figure 4.8: After-cooling unfiltered and filtered ROI coordinate plots for video P-H-RA29M2



(a) MediaPipe landmarks: Frame 284

(b) MediaPipe landmarks: Frame 1138

(c) MediaPipe landmarks: Frame 5979

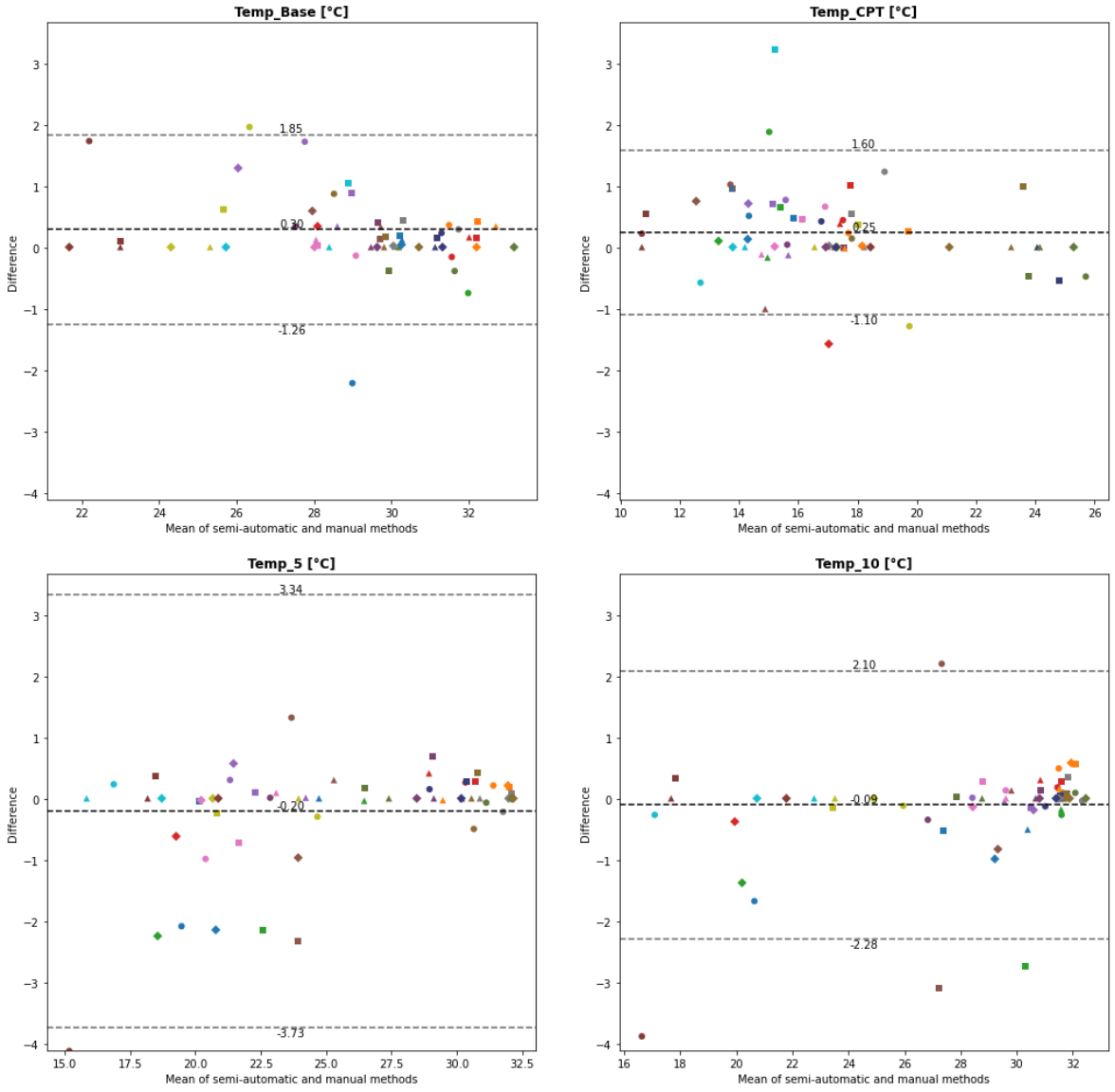
(d) ROIs after filtering: Frame 284

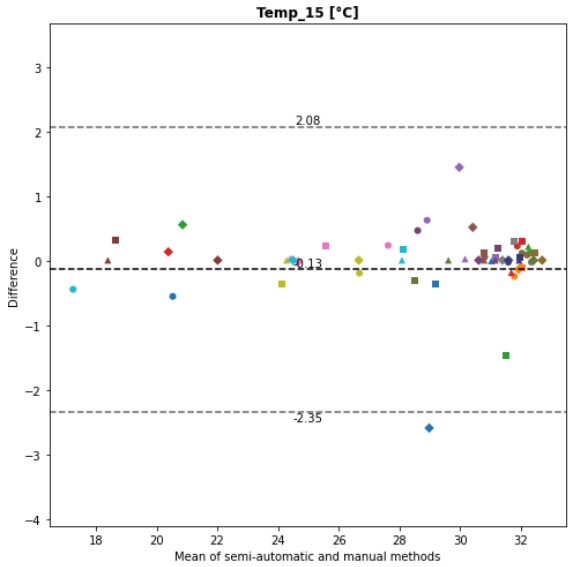
(e) ROIs after filtering: Frame 1138

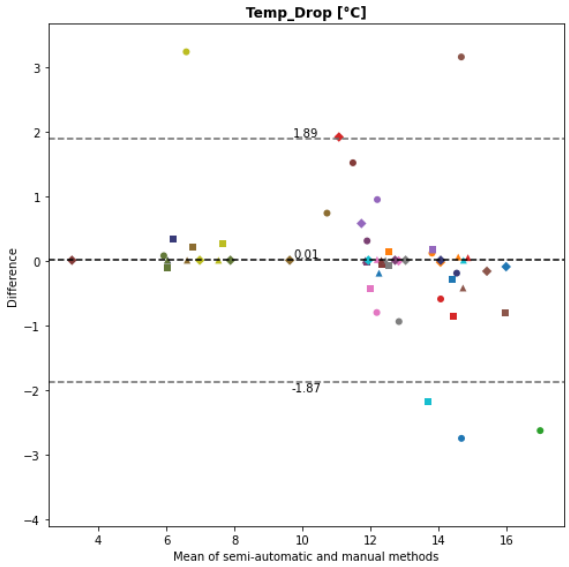
(f) ROIs after filtering: Frame 5979

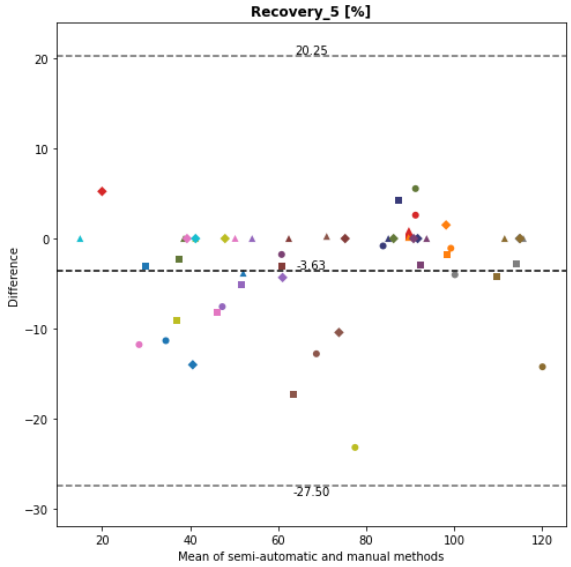
Figure 4.9: Visual comparison between the MediaPipe landmarks and the final ROI locations at three moments in video P-H-RA29M2

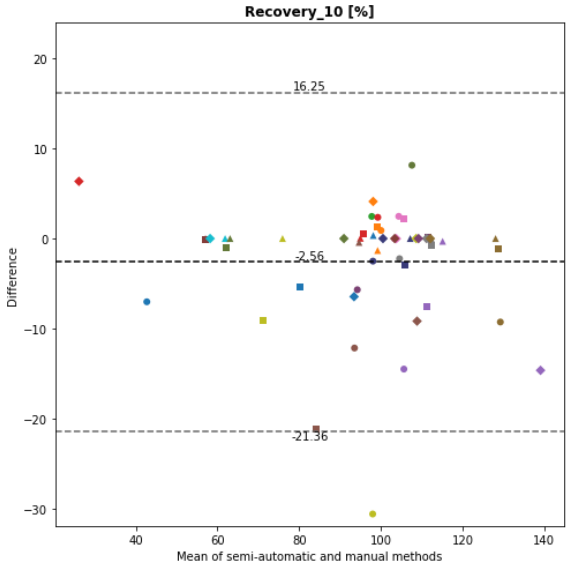
For the six temperature variables, the mean differences all lie between -0.20°C and 0.30°C, suggesting little systemic bias. The distance towards the upper and lower limits varies between 1.35 and 3.53 degrees. The recovery percentage variables Recovery_5/10/15 have mean differences of -3.63%, -2.56% and -3,61% with limits of 23.87%, 18.80% and 25.42% below and above the mean, respectively. t_50 has a lower limit of -2.87, a mean difference of 0.50 and an upper limit of 3.87 minutes. For t_95, these values contribute -2.24, 0.30 and 2.84 minutes from the lower limit to the mean difference and the upper limit. These varying distances to the LoAs indicate considerable variation in the measurements of the two approaches. The scatter dots, representing the differences for individual measurements against the average, are, in most cases, randomly distributed around the mean line without a trend being observed. For the variables Temp_5/10/15 and Recovery_5/10/15, the slight majority of the points lie above the mean difference line, indicating that the semi-automatic method tends to overestimate compared to the manual method. For t_50 and t_95, the opposite is true, with points lying below the mean difference line, indicating that the semi-automatic method tends to underestimate compared to the manual method. For some variables, the LoAs are exceeded, with six ROIs from different videos in the most extreme case (Temp_Drop). Conclusively, apart from a couple of outliers and the observed variance, the methods do not present proportional bias based on the Bland-Altman analysis of the eleven variables.

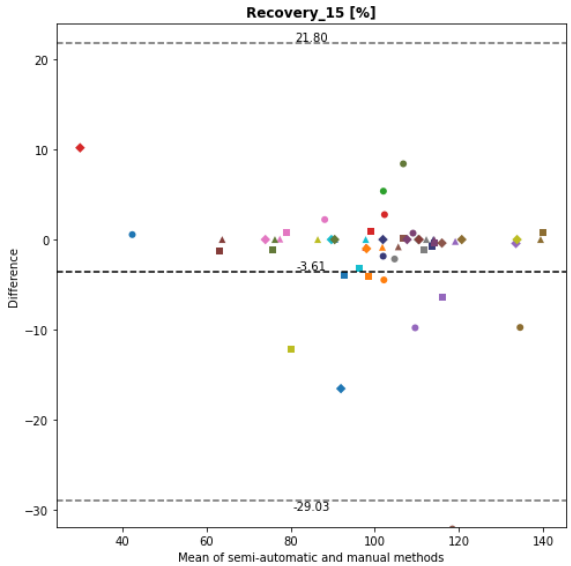


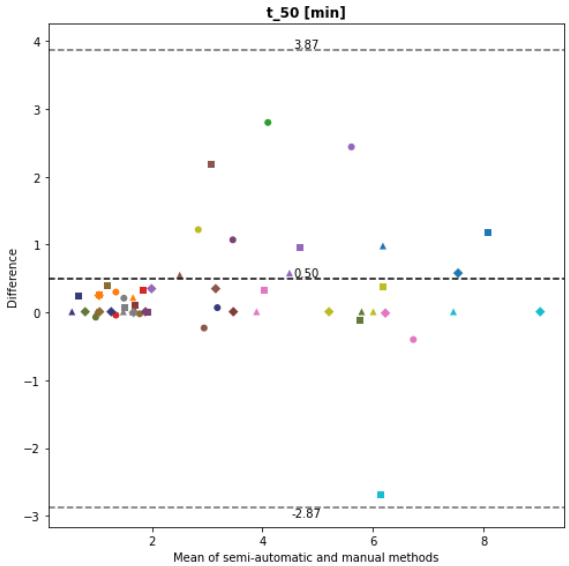

Temp_15 [°C]
2.08
0.13
-2.35
Difference
Mean of semi-automatic and manual methods


Temp_Drop [°C]
1.89
0.01
-1.87
Difference
Mean of semi-automatic and manual methods


Recovery_5 [%]
20.25
-3.63
-27.50
Difference
Mean of semi-automatic and manual methods


Recovery_10 [%]
16.25
-2.56
-21.36
Difference
Mean of semi-automatic and manual methods


Recovery_15 [%]
21.80
-3.61
-29.03
Difference
Mean of semi-automatic and manual methods


t_50 [min]
3.87
0.50
-2.87
Difference
Mean of semi-automatic and manual methods

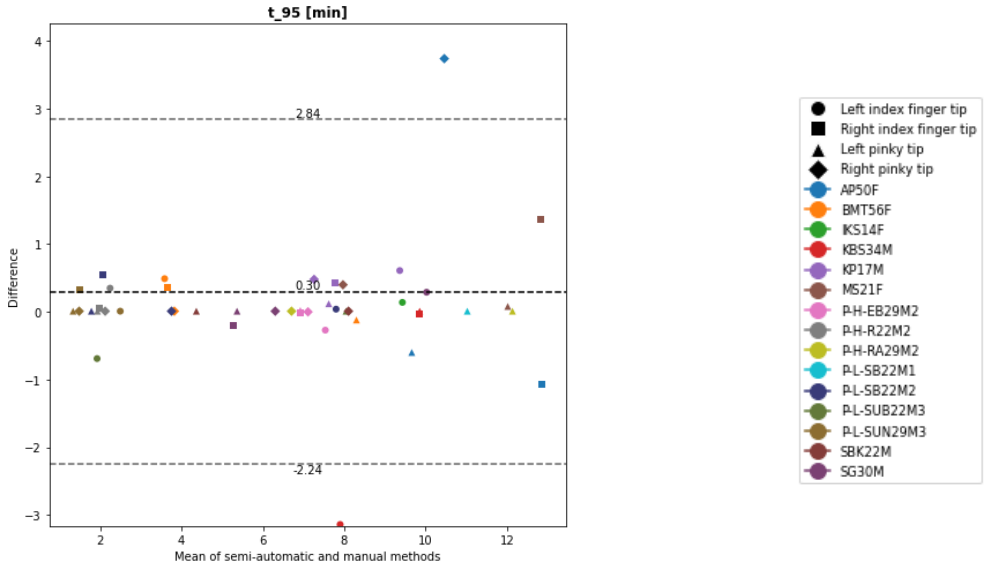


Table 4.5: Bland-Altman plots for the left and right index finger and pinky tips: comparing the semi-automatic and manual methods

Researcher I. Schemkes examined the Bland-Altman plots and selected two videos for visual support: AP50F (blue), which exceeded the limits of agreement for the variables Temp_Base, Temp_15, Temp_Drop and t_95, and BMT56F (orange), which remained relatively close to the mean difference line for across all variables. Three example frames for both the manual method and the semi-automatic approach are illustrated in Figures 4.10 and 4.11, respectively. In the first row of Figure 4.10, it can be observed that the manually placed ROIs of video AP50F are in the correct areas, except for the pinky tips, which slightly extend beyond the finger in frame 9649 (4.10b) and 12953 (4.10c). In subfigures d, e, and f, for the semi-automatic method, it is evident that the left hand is incorrectly annotated. At the beginning of the video, the index finger tip and pinky ROIs are positioned too far to the side of the finger, and at later moments, only the palm is correctly placed. For the video BMT56F, the only inaccuracy, as far as the eye can detect based on Figure 4.11, is the right thumb tips in subfigures b and e. The comparable annotation between the two approaches confirms the similarity in the Bland-Altman plots.

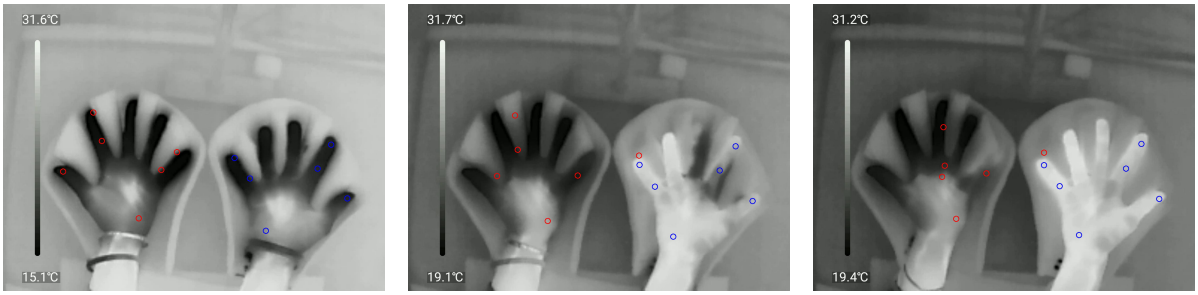
4.3.5. Experiment set-up comparison

Three different experimental setups were compared in terms of MediaPipe tracking performance and the accuracy of ROI locations. These setup variations include a classic arrangement without foam in the hand mould, an experiment setup without the hands being fixated, and the latter setup but with a mirror placed beneath the hands instead of the usual table. From now on, these setups will be referred to as setups A, B, and C, respectively.

Table 4.6 displays the percentage of frames for which MediaPipe was not able to identify landmark coordinates for the left and right hands, during both the before and after cooling stages. Mean values and standard deviations were calculated for the four metrics across the variants. On average, the share of before-cooling frames for which no MediaPipe landmarks could be tracked did not improve in the setup without foam in the mould, compared to the original setup (Table C.1). The performance after cooling even deteriorated in the new setup, from 17% to 35% for the left hand and from 17% to 22% for the right hand. For the videos from set-up B (n = 3) and C (n = 2), the majority of the before-cooling frames and half of the after-cooling frames had all coordinates tracked. The highest percentages of frames with untracked landmarks for set-up B are 28% of frames for the right hand before cooling, and 22% for the left hand after cooling. For experiment C, the pre-CPT stands out with 91% of frames where MediaPipe did not store any right-hand coordinates. Considerable variation is present in the outcomes, similar to performance reports of the classical experimental setups.



(a) Manual approach: Frame 1593 **(b)** Manual approach: Frame 7847 **(c)** Manual approach: Frame 11151

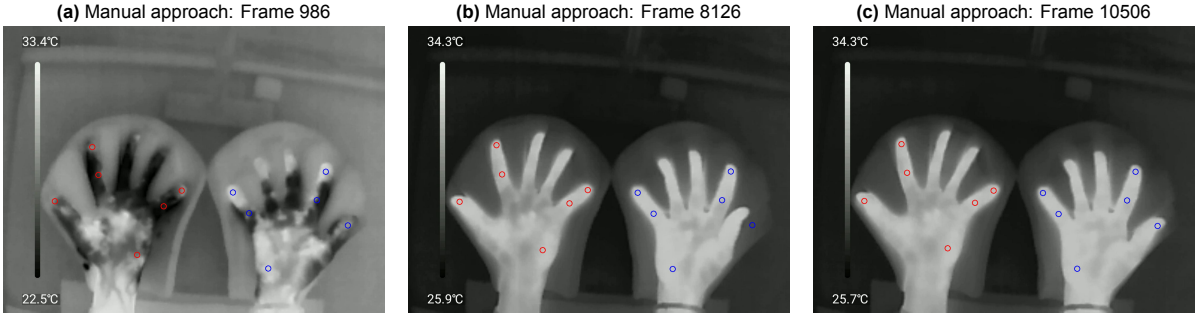


(d) Semi-automatic approach: Frame 1561 **(e)** Semi-automatic approach: Frame 7805 **(f)** Semi-automatic approach: Frame 11187

Figure 4.10: Visual comparison between the semi-automatic and manual method at three moments in video AP50F



(a) Manual approach: Frame 986 **(b)** Manual approach: Frame 8126 **(c)** Manual approach: Frame 10506



(d) Semi-automatic approach: Frame 976 **(e)** Semi-automatic approach: Frame 8053 **(f)** Semi-automatic approach: Frame 10493

Figure 4.11: Visual comparison between the semi-automatic and manual method at three moments in video BMT56F

Table 4.6: MediaPipe tracking performance for three different experimental set-ups

Set-up	File name	Before-cooling frames where left hand coordinates could not be tracked [%]	Before-cooling frames where right hand coordinates could not be tracked [%]	After-cooling frames where left hand coordinates could not be tracked [%]	After-cooling frames where right hand coordinates could not be tracked [%]
A	AN_11F_1	2	98	27	20
	AN_11F_2	0	0	0	19
	BG_44M_1	0	30	44	53
	BG_44M_2	4	0	100	34
	BG_44M_3	0	3	7	27
	DCT_30F_1	100	0	46	0
	DCT_30F_2	0	0	19	1
	Mean ± Std. Deviation	15 ± 37	19 ± 37	35 ± 34	22 ± 18
B	P_H_LIS26F	0	0	22	1
	P_H_MAR24F	0	0	9	0
	P_L_GAN45M	0	28	0	0
	Mean ± Std. Deviation	0 ± 0	9 ± 16	11 ± 11	0 ± 0
C	HD_SP23M	0	0	3	0
	HP_BM23M	0	91	0	1
	Mean ± Std. Deviation	0 ± 0	46 ± 65	2 ± 2	0 ± 1

Set-up:
A) Fixation mould is used without foam attached to it
B) No fixation mould is used
C) No fixation mould is used and a mirror is positioned underneath the hands

Researcher I. Schemkes conducted a visual examination to assess the accuracy of the final ROI locations, i.e., after the coordinates had been filtered. For each video and every ROI, the percentage of after-cooling frames with correctly placed ROIs is presented in Table 4.7. Across all ROIs, set-up A achieves an average correct localisation in 55 ± 14% of the frames. For set-ups B and C, the averages are 64 ± 14% and 82 ± 16% of the frames, respectively. Only the experiment recordings with a mirror background scored higher than the average of 75% for the regular videos.

Table 4.7: Percentage of correctly placed ROIs in after-cooling frames for three different experimental set-ups, based on visual examination

Set-up	File name	Left thumb tip	Left index finger tip	Left index finger pip	Left pinky tip	Left pinky pip	Left palm	Right thumb tip	Right index finger tip	Right index finger pip	Right pinky tip	Right pinky pip	Right palm	Mean ± Std. Deviation
A	AN_11F_1	42	7	11	16	20	13	0	0	0	0	0	0	9 ± 13
	AN_11F_2	87	67	98	27	29	96	60	47	93	49	93	100	70 ± 28
	BG_44M_1	89	78	98	33	62	100	0	0	0	0	0	0	38 ± 44
	BG_44M_2	0	0	0	0	0	0	69	100	100	0	62	100	36 ± 46
	BG_44M_3	96	67	89	80	73	100	96	96	100	0	27	100	77 ± 32
	DCT_30F_1	87	82	98	82	89	98	29	33	40	16	18	36	59 ± 33
	DCT_30F_2	96	93	98	89	100	100	100	100	100	87	100	100	97 ± 5
	Mean ± Std. Deviation	71 ± 36	56 ± 37	70 ± 44	47 ± 36	53 ± 38	72 ± 45	50 ± 42	54 ± 45	62 ± 47	22 ± 34	43 ± 42	62 ± 49	55 ± 14
B	P_H_LIS26F	0	0	0	0	0	0	18	0	93	36	42	42	19 ± 39
	P_H_MAR24F	98	84	100	98	100	100	89	100	100	98	100	100	97 ± 5
	P_L_GAN45M	82	76	93	47	49	100	100	100	100	33	56	100	78 ± 25
	Mean ± Std. Deviation	60 ± 53	53 ± 46	64 ± 56	48 ± 49	50 ± 50	67 ± 58	69 ± 45	67 ± 58	98 ± 4	56 ± 37	66 ± 30	81 ± 33	65 ± 14
C	HD_SP23M	98	100	100	22	36	93	84	91	96	89	100	98	84 ± 26
	HP_BM23M	71	62	78	60	89	93	76	76	89	73	87	100	79 ± 12
	Mean ± Std. Deviation	84 ± 19	81 ± 27	89 ± 16	41 ± 27	62 ± 38	93 ± 0	80 ± 6	83 ± 11	92 ± 5	81 ± 11	93 ± 9	99 ± 2	82 ± 16

Set-up:
A) Fixation mould is used without foam attached to it
B) No fixation mould is used
C) No fixation mould is used and a mirror is positioned underneath the hands

5 Discussion

5.1. Key results

In this study, a method for semi-automatic temperature analysis based on real-time landmark tracking in infrared videos is developed. Determining the temperature within the annotated regions of interest aims to support the TAIL research into the potential of IRT for leprosy diagnosis. The approach is validated using a set of experiment recordings in which the warming up and recovery of the hands after a cold impulse during the CPT is measured.

The model output and the outcomes of various validation evaluations were gathered and analysed. The most important finding from this analysis is the high impact of hand landmark accuracy by MediaPipe. This machine learning task is unable to find coordinates in all frames consistently or tends to annotate landmarks inaccurately in some videos or video segments. This results in a relatively low accuracy for ROI annotation based on visual examination. Attempts were made to correct inaccurate or unknown coordinates with filtering techniques, but in case of substantial distortions due to incorrect landmark input, the model does not consistently succeed. Furthermore, the quantitative outcomes of MediaPipe performance metrics, i.e., the number of frames in which no landmarks were found and the extent of filtering in a video measured by the Euclidean distance, cannot be used as indicators of the quality of ROI localisation.

Applying the semi-automatic model, various variables of interest used within the TAIL research were extracted. Missing hand coordinates before cooling due to failures by MediaPipe can prevent these from being determined. Similarly, incorrect ROI annotations result in unreliable values. Nevertheless, the Bland-Altman analysis indicates that, apart from a few outliers, the outcomes of these variables do not significantly differ from the manual method. Alternative experimental setups were tested for potential improvements in combination with the semi-automatic method. The best results were seen in the setup variant without the fixation mould and with a mirror positioned beneath the hands. However, no premature conclusions can be drawn due to the small sample size. Lastly, it can be stated with certainty that the model outcomes are significantly consistent after repeatable measurements, indicating excellent reliability.

5.2. Results analysis

5.2.1. Output overview

An overview of the model output was generated through the presentation of performance reports, categorised into MediaPipe landmark tracking, temperature extraction, and model running time, alongside the extracted variables of interest.

MediaPipe hand landmark tracking

In the performance reports related to MediaPipe tracking performance, it was observed that landmarks were more frequently completely identified in the before-cooling phase than after-cooling. Despite the presence of extreme variation, the averages for the number of untracked frames were also slightly lower.

One explanation for this is the temperature difference between the hands and the background, which directly affects the video's contrast. Since the difference between a body temperature of approximately 37°C and the cooled room at about 24°C is relatively large, the hand contours are easier to detect in the video. After applying the CPT, the temperature of the hands drops closer to the ambient temperature, making the hands less distinguishable from the background. Additionally, the after-cooling video contains about ten times more frames, increasing the likelihood of complicated frames and thereby reducing the chances of all landmarks being tracked. However, there are a few exceptions to this declaration where the landmarks were not partially or fully tracked before cooling.

Temperature exraction
The percentages of frames in which no minimum or maximum temperature could be extracted are relatively low and, therefore, not considered a significant problem. Given the number of data points and the relatively slow process of the hands warming up over a period of fifteen minutes, this impact remains minimal. The cause of unknown temperature values in the complete before-cooling phase in some instances is the absence of landmark coordinates from one of the hands. Typically, when landmark coordinates are partially missing, linear interpolation can be employed to reconstruct the ROI locations and determine the temperature. However, this method was not viable in these segments due to a complete failure in tracking and, therefore, a lack of input data. Due to the greater contrast change during the recording of hand warming, the variation in this segment is larger. Furthermore, as previously mentioned, the video segment contains about ten times as many frames. These reasons increase the likelihood that MediaPipe can detect the hands at some point, allowing part of the coordinates to be known and the rest to be supplemented with interpolation. Consequently, unknown temperature values must be caused by missing minimum and maximum temperatures. It is important to note that a known temperature does not guarantee its correctness. Especially if many frames have their ROI determined by filtering and interpolation, it is understandable that this can be prone to errors.

Model running time
During the development and training phase, the landmark tracking by MediaPipe has been identified as the most time-consuming part of the process. Consequently, the duration it takes to complete the model can be related to the performance of this component. In videos where MediaPipe is less effective at extracting landmark locations, a relatively longer running time has been observed. Since the total time is determined and not calculated separately for each component of the code, this conclusion cannot be stated with certainty.

Variables of interest
Not all variables could be determined in every case by the developed semi-automatic method. There are two principal reasons why a variable might be unknown and, therefore, appear as NaN in the results. Since a baseline measurement is required, the inability to determine the before-cooling temperature results in an unknown temperature drop and recovery-related variable results. For videos where this occurs in half of the ROIs, the input coordinates of one hand are likely missing. Additionally, it may occur that the 50% or 95% temperature recovery is not achieved within the time of recording. This logically happens more often for 95% recovery since this threshold is higher, confirmed by the lower amount of extracted time values.

5.2.2. Repeatability

The test-retest reliability analysis was conducted to determine the consistency of the results when the model was applied three times to the same data. Utilizing repeatability coefficients, it was assessed whether the MediaPipe task performed consistently on identical input videos before and after cooling. The results of zero indicate perfect repeatability of the automatic part of the developed approach.

Intraclass correlation coefficients were used to assess the intra-observer variability of the manually performed phase selection. The user must indicate the start and end frames of the before and after cooling stages using a trackbar, making the method semi-automatic rather than fully automatic. The high ICC values suggest excellent reliability, supporting the conclusion that intra-observer variability is negligible and that differences in outcomes are due to other variations in input data.

5.2.3. Visual examination

A set with classical experiment recordings was visually examined on ROI accuracy to provide insight into the model's reliability. In some instances, a poor visual score can clearly be attributed to MediaPipe's performance, i.e., the proportion of frames in which no landmarks were found. However, in other videos, there is no cause-and-effect relationship observed, making the metric unsuitable for predicting the ROI accuracy outcome. The issue lies in the quality of the tracking and further processing, which is not measured within the semi-automatic method. The palms on both hands exhibit the highest percentages among the six ROIs, which can be attributed to the larger palm zone compared to the fingers, allowing for a greater margin of error. A possible explanation for the lower accuracy observed for the pinkies could be the smaller target size, where a circle of the same radius is more likely to extend beyond the correct area. The choice to use regions with a radius of ten, aimed at, together with median temperature calculation, ensuring robust temperature determination where the impact of variation within the fingers is minimised, may actually work against achieving accurate results here.

With completely missing regions in some cases before cooling and an average performance after cooling that exhibits significant variation among the videos, the results of this visual examination do not appear promising. Although this sample evaluation is based on a small subset of frames, representing approximately less than one per cent of the total video, it does not suggest good accuracy in ROI localisation.

5.2.4. Effect of coordinate filtering

The Euclidean distances were used to indicate the amount of change after coordinate filtering. Having analysed the results, which include a detailed examination of two examples, the following conclusions can be drawn: firstly, the Euclidean distance between the MediaPipe landmarks and ROI locations after filtering cannot be used as a quality measure for annotation accuracy. In other words, the amount of filtering in a video does not correlate with accurately localised regions on the hand, based on visual examination. Moreover, it is not the quantity of coordinates tracked by MediaPipe, but their quality that is crucial for a good outcome. Videos with many disturbances, i.e., many incorrect landmark locations or excessive variation, yield poor results after coordinate filtering. Therefore, correct landmark tracking performed by MediaPipe is essential to accurately determining the temperature at the correct locations on the hands.

5.2.5. Comparison to the manual method

The Bland-Altman analysis conducted in this study provides a comprehensive comparison between the variables of interest of the semi-automatic and manual methods. These results indicate that, aside from the variance among measurements and a few outliers, these outcomes do not significantly differ from each other. However, since the manual approach does not serve as perfect ground truth and the variables alone do not adequately reflect model performance, the outcomes of the developed method cannot be considered fully reliable.

Slightly higher temperatures from the semi-automatic compared to the manual method were observed. Consequently, the recovery percentages at these time points were also somewhat higher, while the time to 50% and 95% recovery is lower. This is expected given that the calculation of these values relies on the temperature values. One possible explanation for this could be that the temperature measured in the frames around the hands is higher during the experiment, resulting in a higher extracted temperature for an incorrectly placed ROI. The fixation mould, seen as bright circles around the hands, seems to interfere and indicates a higher temperature than the broader surroundings. This phenomenon occurs in Figure 4.10f. However, this would be illogical for segments later in time, as it is expected that the temperature will eventually rise above the constant ambient temperature of 24°C. The latter explanation is supported by the example frames in Figures 4.11e and 4.11, where the hands appear brighter than the surrounding circular shape.

5.2.6. Experiment set-up comparison

Three alternative test setups were explored in an attempt to improve the contrast in the infrared video and consequently enhance MediaPipe tracking performance. For the setups without hand fixation, a significant reduction in the average percentage of frames without extracted landmarks was observed. Since it was previously established that the quality, rather than the quantity, of the annotated landmarks

indicates outcome reliability, the performed visual examination is considered more important. On average, only the setup variant with a mirror background achieved an increase in ROI accuracy score. Due to the small sample size used for these tests, no definitive conclusions can be drawn, and further validation on more data is recommended. The extremely low outliers in the other two experimental setups — table background and either hands in a mould without foam or without fixation — lower the reliability for practical use.

5.3. Model requirements

A list of requirements for the intended model was established at the beginning of the research. The criteria can now be retrospectively reviewed and assessed to determine the extent to which they have been achieved.

1. The model should require as few as possible and at least fewer manual steps than the existing manual method.
2. The model should be at least as reliable and accurate as the manual approach.
3. No or as few additional expenses as possible should be incurred compared to the current process.
4. The model should be easily implemented in the existing research environment.
5. The model should be user-friendly for individuals without prior knowledge of the software.
6. The model should not be bound by location restrictions.
7. The model should be adaptable to accommodate alterations, such as when the input data or the intended output changes.
8. The execution time of the model should remain within reasonable margins.

Since the user only selects the start and end points of the before and after cooling periods manually, the first criterion has been met. Unlike the manual approach, it is not necessary to annotate the ROIs. The accuracy of the designed method compared to the existing manual approach has been analysed, concluding significant similarity for the variables outcomes of both approaches. Items 3, 4, and 6 do not change relative to the current research setting, where the free Python software is also used. The model is user-friendly due to the executable file format, and the underlying code is provided for necessary adjustments (items 5 and 7). The average execution time is 42 minutes, which can be considered acceptable.

5.4. Limitations

In addition to the disadvantages already mentioned in the analysis of the results, there are some limitations of this study that necessitate discussion. These shortcomings relate to MediaPipe hand landmark tracking, coordinate filtering, and other methodological considerations.

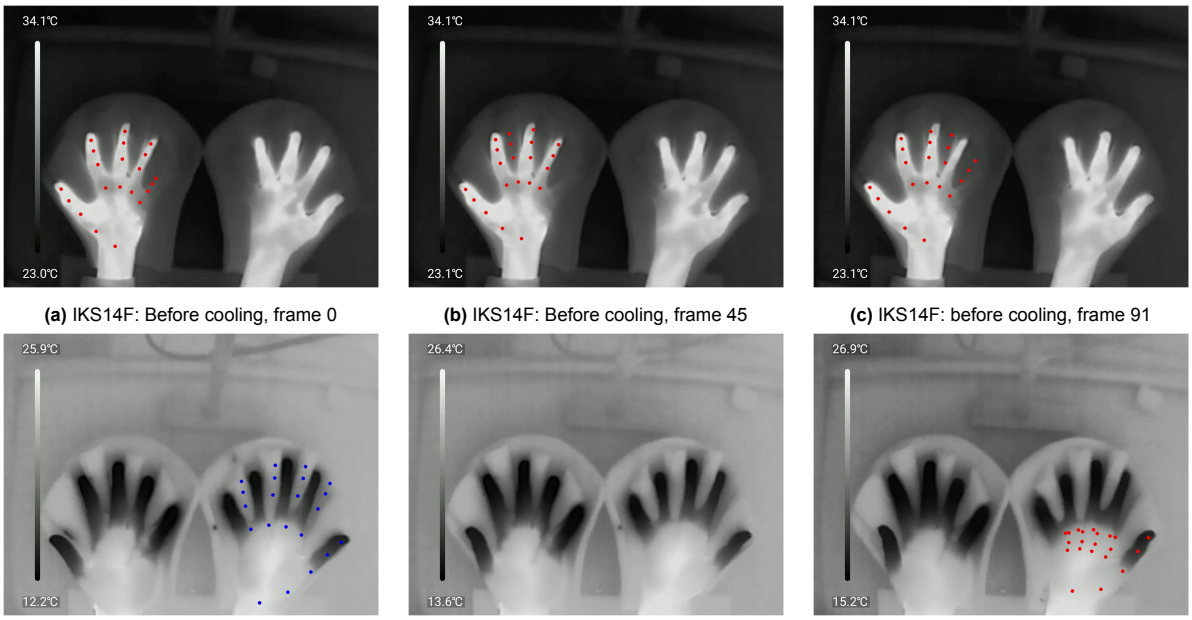
5.4.1. MediaPipe hand landmark tracking

The key element of this model is the tracking of ROIs in an infrared video to provide insight into the response of the peripheral nervous system after the CPT. The more precisely the regions can be determined and followed during rewarming, the more accurate the determined temperature values over time are. The extracted variables of interest concerning temperature and recovery measures must be reliable to support the TAIL research. Since the hands, even in the mould, cannot remain perfectly still for at least fifteen minutes, it is preferable to perform this tracking real-time. The MediaPipe hand landmark tracking task was chosen as the basis for the automatic approach. The potential of this machine-learning model can be substantial for achieving this goal. However, it did not always work well in this context, resulting in incorrectly annotated ROIs and method outcomes being insufficiently reliable.

The first explanation for Mediapipe's failure is that the infrared videos are not similar to the training data. Given the small dataset that was initially available for this study versus the 30,000 images on which the MediaPipe model was trained, this seemed to be the best option for real-time ROI tracking. Additionally, the poor contrast and unclear finger edges in video frames are a problem for landmark localisation. Temperature differences within the hand create lighter or darker areas in the infrared

frames. This is, for example, visible with the pinky finger appearing darker in the first row of Figure 5.1. As a result, the right hand is completely unrecognised, evidenced by the absence of blue dots. In the first frame Subfigure 5.1a, the pinky on the left is recognised but is underestimated in length. Over time landmarks are placed somewhere between the fingers (Figure 5.1b and 5.1c). In the after-cooling frames of a second example, the palm is warmer and thus appears lighter (Figure 5.1, second row). The left hand is not detected in the majority of the video. In addition, an odd hand shape is mistakenly annotated on the right-hand palm, labelled as the left hand (5.1f). Finally, the fixation foam — between the fingers — is sometimes considered as a finger, visible in Subfigure 5.1f. One explanation could be that the hands heat up the foam relative to the ambient temperature, causing it not to blend in with the background. As a result, there are more distortions around the hand, making the hand edges harder to detect. The foam in this example also appears in a relatively similar grey value as the palm, which could make it look more like a part of the hand to MediaPipe.

It is challenging that MediaPipe does not specify the exact problems or obstacles encountered in analysing certain video segments or frames. This ”black box” nature complicates identifying appropriate solutions and improving ROI annotation. This study has attempted to develop a measure of performance since MediaPipe does not provide one. Based on the validations, it can be concluded that the proportion of frames with detected landmarks or the Euclidean distance after coordinate filtering cannot predict the accuracy of ROI annotation.



(a) IKS14F: Before cooling, frame 0 **(b)** IKS14F: Before cooling, frame 45 **(c)** IKS14F: before cooling, frame 91

(d) SBK22M: After cooling, frame 3201 **(e)** SBK22M: After cooling, frame 5602 **(f)** SBK22M: After cooling, frame 11205

Figure 5.1: Example frames of MediaPipe's tracking failure: red dots for the left hand and blue dots for the right hand

5.4.2. Coordinate filtering

To compensate for missing or incorrect ROI coordinates, filtering is introduced in this study. Various filtering methods, based on Fourier transforms, variance, and the first and second derivatives, have been tested, but the chosen method was able to approach the actual ROI locations best. It was previously observed that this does not work when the proportion of incorrectly found landmarks or distortions is too large. This, again, highlights the importance of accurate annotation by MediaPipe for the success of the semi-automatic analysis. In some cases coordinate filtering has made a positive difference, i.e., incorrectly annotated MediaPipe landmarks were shifted towards the aimed region. However, it was observed that the concept of real-time tracking disappears as the x and y coordinates are stabilised over time. Following rapid hand movements, which would be a desired improvement compared to the manual analysis, is no longer possible. Additionally, a disadvantage of filtering is that it alters not only outliers but also a portion of the correct data. Assuming that the ROIs are accurately tracked or filtered based on visual examination, variations are still present in the exact circle location on the

hand. Shiftings of the ROIs, both vertically and horizontally within, over time have been observed in the visual examination. This was approved, as long as the circular region remained between the hand edges, but it should be examined more strictly to increase reliability. An inconsistent location where the temperature is determined can affect the calculated outcomes of the variables compared to reality.

5.5. Recommendations for future research

Given the established importance of accurate ROI annotation, the primary recommendation is to improve this into an automatic model that generates reliable results. Correctly localised ROIs could lead to a robust temperature analysis serving the TAIL research objectives. Firstly, it is advised to investigate whether the input data can be modified to improve MediaPipe's ability to locate the hands. A Canny edge detection algorithm has already been applied in an attempt to simplify the more complex frames for landmark tracking. However, this did not improve the results. Extensive research could look into methods of contrast enhancement techniques of the video. Secondly, the use of a different infrared camera - with a higher resolution - or alternative colour settings could be explored for improved detection. Lastly, further research into experiment setup alternatives might be useful.

Alternatively, it is advisable to develop a machine-learning model for automatic landmark tracking in infrared videos by training it on the appropriate data. By training the method on videos that will ultimately be used in this context, specifically white-hot IR videos in the identical experiment setup, the outcomes are likely to improve. For this, more data acquisition might be required, and the influence of low contrast must still be taken into account. In both scenarios, whether using MediaPipe tracking or a newly developed machine learning method, it is desirable for the model to provide feedback on the accuracy of the ROI annotation itself. This feature would allow users, namely researchers from the TAIL initiative, to assess the reliability of the results.

If, after implementing the above suggestions, the analysis remains insufficiently reliable for the TAIL research, transitioning to a hybrid form between automatic and manual analysis could be considered. The developed model is already semi-automatic, but after the initial part of before and after cooling phase selection, no further human intervention is required. A visual assessment of sample frames could be inserted before proceeding with the extraction of temperature values. This would allow a researcher to confirm whether they agree with the ROI locations determined by the algorithm. If this is not the case, manual annotation followed by linear interpolation can serve as an alternative reconstruction method. This approach would take more time and might introduce potential reproducibility differences, but it could potentially improve the outcomes compared to the current situation.

To validate whether a developed method is reliable or approximates reality as closely as possible, a ground truth measurement is essential. For this research, which is still in its initial phase, no baseline is available. The manual method has been used for comparison, but this approach has shown imperfections as well. To compensate for this as effectively as possible, a visual examination of sample frames was used to place the output of quantitative analysis in context. In the future, a reliable ground truth could further improve the quality of the model. This could be established by having one of the researchers manually annotate the twelve ROIs very precisely while going through a video. By doing this for a sufficiently large subset of frames, a ground truth of accurate ROI locations is established. To validate the temperature determination, it is recommended to compare the results with those obtained from another infrared camera. It might be interesting to review the differences between the results from a high-resolution camera and those from the low-cost camera chosen for the TAIL research initiative. To measure the response of the peripheral nervous system after CPT, which is the focus of the experiment, it is crucial that the IRT accurately captures the temperature difference over time.

6

Conclusion

This study aimed to develop and validate a semi-automatic temperature analysis method using real-time hand landmark tracking in infrared videos. The goal was to enhance the accuracy and efficiency of temperature analysis for diagnosing early-stage leprosy, thereby addressing the limitations of the existing manual method and supporting the research initiative: Thermographic Assessment of Autonomic Impairment in Leprosy

A semi-automatic Python model was developed, integrating MediaPipe for hand landmark tracking with custom temperature extraction and ROI localisation algorithms. The model successfully reduced the manual annotation workload, offering a more efficient approach than the manual method. The repeatability analyses indicated that the model's automatic component consistently produced reliable results with negligible intra-observer variability. Visual examination revealed that while the model performed well in many cases, its accuracy heavily depended on the quality of hand landmark detection by MediaPipe. Inaccurate landmark tracking led to significant errors in ROI localisation and temperature determination.

The Bland-Altman analysis demonstrated that the semi-automatic method's results generally agreed with the manual method, although some discrepancies existed, particularly in cases of poor landmark detection. Various experimental setups were tested to improve tracking performance. The use of a mirror background showed some promise in increasing ROI accuracy, but further validation with a larger sample size is necessary.

Improving hand landmark detection in infrared videos is crucial for future research. Future research should focus on enhancing the performance of MediaPipe or developing a custom machine-learning model robust enough to track landmarks accurately in low-contrast frames and across different hand types. Implementing a feedback mechanism within the model to indicate the accuracy of ROI annotation could enhance reliability, allowing researchers to assess the trustworthiness of the results before interpretation and use in the TAIL research. Further testing with different experimental setups, including varying backgrounds and fixation methods, should be conducted to ensure minimal interference from surrounding objects and optimise contrast in infrared videos, thereby enhancing landmark detection and overall model performance.

The development of a semi-automatic temperature analysis model marks a step forward in utilising infrared thermography for early-stage leprosy diagnosis. While the current model shows promise, its success depends too much on MediaPipe's hand landmark detection accuracy. Future enhancements in the model or infrared data, combined with thorough validation, are essential to realise the potential of supporting the TAIL research.

References

[1] Kou-Huang Chen et al. “Leprosy: a review of epidemiology, clinical diagnosis, and management”. In: *Journal of Tropical Medicine* (2022).

[2] World Health Organization. “Global leprosy (Hansen disease) update, 2020: impact of COVID-19 on global leprosy control”. In: *Wkly Epidemiol Rec* 96.36 (2021), pp. 421–444.

[3] World Health Organization. *Leprosy reactions*. Available online: `https://www.who.int/news-room/fact-sheets/detail/leprosy`. 2023.

[4] W Alemu Belachew and B Naafs. “Position statement: LEPROSY: Diagnosis, treatment and follow-up”. In: *Journal of the European Academy of Dermatology and Venereology* 33.7 (2019), pp. 1205–1213.

[5] Osvaldo JM Nascimento. “Leprosy neuropathy: clinical presentations”. In: *Arquivos de neuro-psiquiatria* 71 (2013), pp. 661–666.

[6] BR Agrawal and RI Agrawal. “Arteriography in leprosy.” In: *Indian journal of leprosy* 57.1 (1985), pp. 138–145.

[7] W Cairns S Smith et al. “Predicting neuropathy and reactions in leprosy at diagnosis and before incident events—results from the INFIR cohort study”. In: *PLoS neglected tropical diseases* 3.8 (2009), e500.

[8] Indira P Kahawita, Stephen L Walker, and Diana NJ Lockwood. “Leprosy type 1 reactions and erythema nodosum leprosum”. In: *Anais brasileiros de dermatologia* 83 (2008), pp. 75–82.

[9] Dhanya Sankaran et al. “Role of histopathology in predicting type 1 lepra reaction in borderline tuberculoid leprosy”. In: *Indian dermatology online journal* 11.4 (2020), pp. 586–589.

[10] Eva María Sánchez-Martínez et al. “Erythema nodosum leprosum successfully treated with apremilast: more effective and safer than classic treatments?” In: *International Journal of Dermatology* 60.2 (2021).

[11] Jose Augusto da Costa Nery et al. “Understanding the type 1 reactional state for early diagnosis and treatment: a way to avoid disability in leprosy”. In: *Anais brasileiros de dermatologia* 88 (2013), pp. 787–792.

[12] Samuel L Moschella. “An update on the diagnosis and treatment of leprosy”. In: *Journal of the American Academy of Dermatology* 51.3 (2004), pp. 417–426.

[13] Prof Pierre Couppie Dr Elise Ouedraogo. *Leprosy reactions*. DermNet: All about the skin. Available online: `https://dermnetnz.org/topics/lepra-reactions`. 2017.

[14] World Health Organization et al. *WHO Expert Committee on leprosy: eighth report*. World Health Organization, 2012.

[15] Bangladesh Nilphamari. “Incidence rates of acute nerve function impairment in leprosy: a prospective cohort analysis after 24 months (The Bangladesh Acute Nerve Damage Study)”. In: *Lepr Rev* 71.1 (2000), pp. 8–33.

[16] Jean M Watson et al. “Ballpoint pen testing: light touch versus deep pressure”. In: *Leprosy Review* 73.4 (2002), pp. 392–393.

[17] JA Bell. “Sensibility testing, state of the art”. In: *Rehabilitation of the Hand* (1989), pp. 575–584.

[18] Wim H van Brakel et al. “Early diagnosis of neuropathy in leprosy—comparing diagnostic tests in a large prospective study (the INFIR cohort study)”. In: *PloS Neglected tropical diseases* 2.4 (2008), e212.

[19] Annelies Wilder-Smith and Einar Wilder-Smith. “Electrophysiological evaluation of peripheral autonomic function in leprosy patients, leprosy contacts and controls.” In: *International Journal of Leprosy and Other Mycobacterial Diseases: Official Organ of the International Leprosy Association* 64.4 (1996), pp. 433–440.

[20] Aretusa Lopes Cavalheiro et al. “Thermographic analysis and autonomic response in the hands of patients with leprosy”. In: *Anais Brasileiros de Dermatologia* 91 (2016), pp. 274–283.

[21] Barid Baran Lahiri et al. “Medical applications of infrared thermography: a review”. In: *Infrared physics & technology* 55.4 (2012), pp. 221–235.

[22] Xavier Maldague. “Theory and practice of infrared technology for nondestructive testing”. In: *(No Title)* (2001).

[23] Michael F Modest and Sandip Mazumder. *Radiative heat transfer*. Academic press, 2021.

[24] Neil C Abbot et al. “Cold fingers in leprosy”. In: *International journal of leprosy and other mycobacterial diseases* 60 (1992), pp. 580–580.

[25] A Chandrahasan Johnson et al. “Lower limb angiography in leprosy”. In: *Radiology* 126.2 (1978), pp. 327–332.

[26] PL Wahi et al. “Peripheral arteriographic studies in leprosy”. In: *Clinical Radiology* 27.3 (1976), pp. 365–370.

[27] LJ Jiang et al. “A perspective on medical infrared imaging”. In: *Journal of medical engineering & technology* 29.6 (2005), pp. 257–267.

[28] Subramaniam Bagavathiappan et al. “Infrared thermal imaging for detection of peripheral vascular disorders”. In: *Journal of medical physics/Association of Medical Physicists of India* 34.1 (2009), p. 43.

[29] Jaap J van Netten et al. “Infrared thermal imaging for automated detection of diabetic foot complications”. In: *Journal of diabetes science and technology* 7.5 (2013), pp. 1122–1129.

[30] Qiang Zhou et al. “Early diagnosis of diabetic peripheral neuropathy based on infrared thermal imaging technology”. In: *Diabetes/Metabolism Research and Reviews* 37.7 (2021), e3429.

[31] Liliane Marques de Pinho Tiago et al. “Assessment of neuropathic pain in leprosy patients with relapse or treatment failure by infrared thermography: A cross-sectional study”. In: *PLoS Neglected Tropical Diseases* 15.9 (2021), e0009794.

[32] Suraj Maharjan. *Proposal: Thermographic Assessment of Autonomic Impairment in Leprosy (TAIL project)*. INF’s Green Pastures Hospital and Shining Hospital Banke. 2024.

[33] Jennifer L Wirch et al. “Cold pressor test protocol to evaluate cardiac autonomic function”. In: *Applied physiology, nutrition, and metabolism* 31.3 (2006), pp. 235–243.

[34] Guillaume Lamotte et al. *The expanding role of the cold pressor test: a brief history*. 2021.

[35] Manar Aljabri et al. “Towards a better understanding of annotation tools for medical imaging: a survey”. In: *Multimedia tools and applications* 81.18 (2022), pp. 25877–25911.

[36] Jaspreet Singh and Ajat Shatru Arora. “Automated approaches for ROIs extraction in medical thermography: a review and future directions”. In: *Multimedia Tools and Applications* 79 (2019), pp. 15273–15296.

[37] Jaspreet Singh and Ajat Shatru Arora. “A framework for enhancing the thermographic evaluation on characteristic areas for paranasal sinusitis detection”. In: *Infrared Physics & Technology* 85 (2017), pp. 457–464.

[38] UNI-T. *UTi721M Thermal Camera Specifications*. Available online: `https://thermal.uni-trend.com/product/uti721m/#Specifications9ebf-edd4`.

[39] Python™. *Python version 3.12*. Available online: `https://www.python.org`. 2024.

[40] Mediapipe. *Hand landmarks detection guide*. Google: AI for Developers. Available online: `https://ai.google.dev/edge/mediapipe/solutions/vision/hand_landmarker`. 2024.

[41] Mediapipe. *MediaPipe Hands*. MediaPipe: Read the Docs. Available online: `https://mediapipe.readthedocs.io/en/latest/solutions/hands.html#model_complexity`. 2023.

[42] Steven Walfish. “A review of statistical outlier methods”. In: *Pharmaceutical technology* 30.11 (2006), p. 82.

[43] Jiawei Yang, Susanto Rahardja, and Pasi Fränti. “Outlier detection: how to threshold outlier scores?” In: *Proceedings of the international conference on artificial intelligence, information processing and cloud computing*. 2019, pp. 1–6.

[44] HP Vinutha, B Poornima, and BM Sagar. “Detection of outliers using interquartile range technique from intrusion dataset”. In: *Information and decision sciences: Proceedings of the 6th international conference on ficta*. Springer. 2018, pp. 511–518.

[45] MN Noor et al. “Filling missing data using interpolation methods: Study on the effect of fitting distribution”. In: *Key Engineering Materials* 594 (2014), pp. 889–895.

[46] Hina Magsi et al. “Analysis of signal noise reduction by using filters”. In: *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. IEEE. 2018, pp. 1–6.

[47] Sharmila Vaz et al. “The case for using the repeatability coefficient when calculating test–retest reliability”. In: *PloS one* 8.9 (2013), e73990.

[48] David L Raunig et al. “Quantitative imaging biomarkers: a review of statistical methods for technical performance assessment”. In: *Statistical methods in medical research* 24.1 (2015), pp. 27–67.

[49] Chia-Cheng Chen and Huiman X Barnhart. “Assessing agreement with intraclass correlation coefficient and concordance correlation coefficient for data with repeated measures”. In: *Computational Statistics & Data Analysis* 60 (2013), pp. 132–145.

[50] Terry K Koo and Mae Y Li. “A guideline of selecting and reporting intraclass correlation coefficients for reliability research”. In: *Journal of chiropractic medicine* 15.2 (2016), pp. 155–163.

[51] Per-Erik Danielsson. “Euclidean distance mapping”. In: *Computer Graphics and image processing* 14.3 (1980), pp. 227–248.

[52] Catey Bunce. “Correlation, agreement, and Bland–Altman analysis: statistical analysis of method comparison studies”. In: *American journal of ophthalmology* 148.1 (2009), pp. 4–6.

[53] Shahab Haghayegh et al. “A comprehensive guideline for Bland–Altman and intra class correlation calculations to properly compare two methods of measurement and interpret findings”. In: *Physiological measurement* 41.5 (2020), p. 055012.

A

User guide: Experiment Protocol

Materials

- Approximately 3 litres of cold water (from the fridge, 5°C)
- Hand fixation set-up
- Mobile phone, USB extension cord, infrared camera, thermography app
- (Paper) towel
- Stopwatch for experiment timing
- A spray can with 1% sodium hypochlorite

Preparation before the arrival of the patient

1. Ensure the room air conditioning is set to 24 degrees Celsius.
2. Install the infrared camera in the set-up; use a USB extension cord to connect it to the phone.
3. Start the thermography app; check all settings; check the hand position in the image since both hands should be fully visible in the picture.
4. Ensure the phone is in the correct orientation such that the temperature indicator is beside the hand position and not overlaying.
5. Ensure sufficient cold water is available in the fridge. Do not get water out of the fridge yet.
6. Ensure the two canisters are available and ready to be filled with cold water during the experiment. Do not fill them with water yet.

Data collection

Welcome and acclimatization in the room

1. Welcome the patient
2. Go through an informed consent form and have it signed.
3. Assign a unique patient identifier on the data collection form.
4. After 15 minutes of acclimatization in the room, proceed to the Data collection.

Pre-test data collection

1. Position the participant in the chair.
2. Take initial measurements and record on data collection form:
    - room temperature (Digital Indoor Thermo-Hygrometer),
    - room humidity (Digital Indoor Thermo-Hygrometer),

- axillary body temperature (digital thermometer),
- blood pressure through upper measurement (Aneroid sphygmomanometer),
- Pulse rate (pulse oximeter through toes).

Cold Pressor Test

1. Position the hands in the hand fixation set-up comfortably.
2. Mark the hand positions in the set-up using the soft filler pieces.
3. Instruct the participant to keep hands and fingers very still; absolutely no movement allowed.
4. Start the infrared video recording in the phone app.
5. Prepare 2 canisters of cold water, fill up to level indication.
6. Measure water temperature and record on a data collection form.
7. Carefully, one by one, remove both hands from the set-up without changing the soft filler pieces.
8. Request participant to submerge both hands up to the wrist in cold water and hold still.
9. Start timer when submersion is up to wrist.
10. After 1 minute, remove hands from cold water and immediately dry the hands using a (paper) towel. Pay attention to drying the palmar side of the hands, as the remaining water will cause reflections for the infrared camera.
11. Place the hands of the participant back into the set-up, one by one, carefully, to get back to exactly the same hand and finger position as before the cold pressor test.
12. Repeat instruction to the participant to keep hands and fingers very still; absolutely no movement allowed. Indicate that the cold pressor test is to be repeated if movement occurs.
13. Start timer
14. Measure pulse rate using a pulse oximeter on the toe.
15. After 15 minutes, stop the video recording.
16. Check if the video is stored.
17. Participant can remove hands from the set-up.

Post-test data collection (after 15 min)

1. Take measurements and record on data collection form:
    - body temperature,
    - blood pressure,
    - pulse rate.
2. Take pain score as experienced during CPT.
3. Check if all data is collected.

Thank the patient for participation and dismiss the patient

Data storage

1. Change the video filename to the assigned unique patient identifier.
2. Upload the video to the dedicated Google Drive folder that collects new videos.
3. Store the data collection form and signed consent form in the designated folder.
4. Share all collected data with the principal investigator.

Cleanup and disinfect set-up

1. Discard the water from the canisters,
2. Disinfect canisters using sodium hypochlorite.
3. Disinfect arm/hand rests using sodium hypochlorite.
4. Rinse with clean water before the next use.
5. Place water back in the fridge.

Thermographic Assessment of Autonomic Impairment in Leprosy [32]

B

User Guide: Semi-Automatic Temperature Analysis

This user guide explains how to use the model for semi-automatic temperature analysis in infrared videos. A few simple steps will guide you through performing the task.

Installing the model

The model and the accompanying files required for semi-automatic temperature analysis can be obtained from the shared Google Drive Autonomic innervation.

1. Go to the Google Drive Folder: *Autonomic Innervation*, *5 - Automatic IR image - Irene Schemkes*
2. Download the entire folder *Semi-Automatic Temperature Analysis* and save it locally on your computer.
3. This folder should contain the following files and folder:
    - SemiAutomaticModel.exe
    - SemiAutomaticModel.py
    - hand_landmarker.task
    - Folder: *Data*, containing subfolders: *1. Videos*, *2. Processed videos* and *3. Output*
    - User Guide

Running the .exe file

1. Place the videos you want to analyse in the subfolder *1. Videos*
2. Open the SemiAutomaticModel.exe file to start the analysis
3. You will be guided through the code to perform the manual steps:
    - File selection from the *1. Videos* subfolder
    - Selection of the start and end frames of both the before and after cooling phases
4. The rest will be performed automatically. Do not close the window until the model has finished

Running the .py file

The following steps on how to run the Python code are identical to those for the manual method developed by F. van den Bogaert.

1. Ensure you have installed Python and Anaconda or Miniconda

2. Install Tesseract by following the steps on this website:
   `https://tesseract-ocr.github.io/tessdoc/Installation.html`
3. Run the following lines in your terminal or command prompt:
   - Pip install imageio
   - Pip install opencv-python
   - Pip install xlsxwriter
   - Pip install pytesseract
4. Store tesseract.exe on your computer
5. Insert the '*Path to tesseract.exe*' in the Python script
6. Run the Python code

Transfer a .py file to an .exe file

Similarly, the steps explaining the creation of an .exe file are identical to the guide of F. van den Bogaert.

1. Run the following lines in your terminal or command prompt:
   - Pip install pyinstaller
   - cd '*Path to the folder with the Python file you want to transfer*'
   - pyinstaller –onefile *Name Python file*.py
2. The .exe file will now appear in the same folder where the Python file is stored

C

In-Depth Results

Output overview

Performance reports

Table C.1: Performance reports for n = 15 infrared videos recorded according to the classical experiment setup

	1	2	3	4	5	6	7	8	9	10	11	12	13
AP50F	93	4	53	12	0	0	0	0	1	5	0	5	42
BMT56F	0	0	12	2	28	0	0	28	9	3	0	11	31
IKS14F	0	100	62	20	0	0	0	100	1	1	0	1	61
KBS34M	0	0	2	2	0	0	0	0	1	2	0	3	37
KP17M	1	0	30	14	5	0	0	5	2	4	0	5	48
MS21F	0	38	2	23	0	0	0	0	0	2	0	2	28
P-H-EB29M2	0	0	16	3	0	0	0	0	3	0	0	3	31
P-H-R22M2	0	0	0	0	0	0	0	0	1	2	0	2	26
P-H-RA29M2	1	84	4	25	5	0	0	5	14	7	0	19	31
P-L-SB22M1	100	0	14	50	0	0	0	100	0	9	0	10	73
P-L-SB22M2	0	0	0	7	0	0	0	0	1	4	0	5	33
P-L-SUB22M3	2	0	13	1	14	0	0	14	3	3	0	5	36
P-L-SUN29M3	7	0	12	8	0	0	0	0	9	3	0	11	60
SBK22M	0	0	26	78	0	0	0	0	2	26	0	26	57
SG30M	0	0	7	7	0	0	0	0	1	4	0	4	31
Mean ± Std. Dev.	14 ± 34	15 ± 33	17 ± 19	17 ± 21	3 ± 8	0 ± 0	0 ± 0	17 ± 35	3 ± 4	5 ± 6	0 ± 0	7 ± 7	42 ± 15

Labels 1-13:
1) Before cooling frames where left-hand coordinates could not be tracked [%],
2) Before cooling frames where right-hand coordinates could not be tracked [%],
3) After cooling frames where left-hand coordinates could not be tracked [%],
4) After cooling frames where right-hand coordinates could not be tracked [%],
5) Before cooling frames where the minimum temperature could not be read from the scale [%],
6) Before cooling frames where maximum temperature could not be read from the scale [%],
7) Before cooling frames where ROI (radius 10) exceeds the frame boundaries [%],
8) Before cooling frames where ROI temperature could not be determined [%],
9) After cooling frames where the minimum temperature could not be read from the scale [%],
10) After cooling frames where maximum temperature could not be read from the scale [%],
11) After cooling frames where ROI (radius 10) exceeds the frame boundaries [%],
12) After cooling frames where ROI temperature could not be determined [%],
13) Model running time [minutes].

Variables of interest

Table C.2: Percentage of ROIs for which the eleven variables of interest could be determined

	Temp_Base	Temp_CPT	Temp_5	Temp_10	Temp_15	Temp_Drop	Recovery_5	Recovery_10	Recovery_15	t_50	t_95
AP50F	100	100	100	100	100	100	100	100	100	92	42
BMT56F	100	100	100	100	100	100	100	100	100	100	100
IKS14F	50	100	100	100	100	50	50	50	50	33	33
KBS34M	100	100	100	100	100	100	100	100	100	83	67
KP17M	100	100	100	100	100	100	100	100	100	100	92
MS21F	100	100	100	100	100	100	100	100	100	100	83
P-H-EB29M2	100	100	100	100	100	100	100	100	100	100	83
P-H-R22M2	100	100	100	100	100	100	100	100	100	100	100
P-H-RA29M2	100	100	100	100	100	100	100	100	100	83	58
P-L-SB22M1	50	100	100	100	100	50	50	50	50	50	25
P-L-SB22M2	100	100	100	100	100	100	100	100	100	83	67
P-L-SUB22M3	100	100	100	100	100	100	100	100	100	100	33
P-L-SUN29M3	100	100	100	100	100	100	100	100	100	100	100
SBK22M	100	100	100	100	100	100	100	100	100	83	25
SG30M	100	100	100	100	100	100	100	100	100	100	92
Mean ± Std. Dev.	93 ± 18	100 ± 0	100 ± 0	100 ± 0	100 ± 0	93 ± 18	93 ± 18	93 ± 18	93 ± 18	87 ± 20	67 ± 29

Example videos

P-H-R22M2

Table C.3: Extracted variables of interest of video P-H-R22M2

	Left thumb tip	Right thumb tip	**Left index finger tip**	**Right index finger tip**	Left index finger pip	Right index finger pip	**Left pinky tip**	**Right pinky tip**	Left pinky pip	Right pinky pip	Left palm	Right palm
Temp_Base [C]	30.66	30.94	**31.59**	**30.08**	31.29	31.44	**28.78**	**30.04**	30.57	31.21	31.15	31.62
Temp_CPT [C]	20.35	21.09	**18.29**	**17.52**	22.44	21.44	**17.13**	**17.01**	18.91	18.38	27.23	27.23
Temp_5 [C]	31.63	32.23	**31.87**	**32.04**	30.20	31.61	**30.04**	**31.96**	29.01	31.52	31.97	32.25
Temp_10 [C]	31.82	31.98	**32.35**	**31.67**	31.89	32.19	**30.97**	**31.48**	31.59	31.8	32.09	32.32
Temp_15 [C]	31.87	31.94	**32.38**	**31.63**	31.77	32.25	**30.33**	**31.38**	31.49	31.72	32.16	32.33
Temp_Drop [C]	10.32	9.85	**13.30**	**12.55**	8.85	10.01	**11.66**	**13.03**	11.67	12.83	3.92	4.40
Recovery_5 [%]	109.41	113.12	**102.09**	**115.58**	87.76	101.64	**110.8**	**114.75**	86.65	102.38	121.02	114.35
Recovery_10 [%]	111.21	110.56	**105.73**	**112.65**	106.87	107.43	**118.75**	**111.08**	108.69	104.58	123.99	115.81
Recovery_15 [%]	111.70	110.20	**105.91**	**112.38**	105.49	108.10	**113.27**	**110.25**	107.90	103.92	125.88	116.12
t_50 [min]	1.19	1.03	**1.39**	**1.47**	2.15	1.90	**1.86**	**1.69**	2.70	1.99	1.15	1.05
t_95 [min]	1.76	1.76	**2.08**	**1.94**	5.93	3.72	**2.75**	**2.13**	6.32	3.38	1.88	2.06

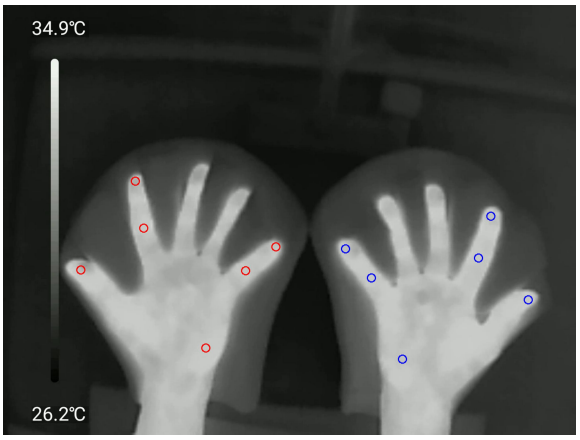
AP50F

Table C.4: Extracted variables of interest of video AP50F

	Left thumb tip	Right thumb tip	**Left index finger tip**	**Right index finger tip**	Left index finger pip	Right index finger pip	**Left pinky tip**	**Right pinky tip**	Left pinky pip	Right pinky pip	Left palm	Right palm
Temp_Base [C]	30.68	30.63	**30.10**	**30.13**	30.22	30.12	**30.82**	**30.24**	30.17	30.88	31.01	30.76
Temp_CPT [C]	14.97	16.38	**14.07**	**15.59**	15.19	17.14	**13.65**	**14.22**	14.16	15.31	19.04	20.53
Temp_5 [C]	19.19	19.93	**20.51**	**20.17**	18.76	19.59	**22.27**	**21.85**	18.75	20.71	24.94	28.08
Temp_10 [C]	20.41	27.99	**21.48**	**27.66**	21.88	23.81	**29.02**	**29.71**	20.64	29.26	27.45	30.11
Temp_15 [C]	25.42	29.94	**20.81**	**29.37**	22.31	28.80	**28.00**	**30.28**	24.55	30.60	27.02	30.31
Temp_Drop [C]	15.71	14.24	**16.04**	**14.54**	15.03	12.98	**17.17**	**16.02**	16.02	15.57	11.96	10.23
Recovery_5 [%]	26.85	24.92	**40.16**	**31.51**	23.78	18.84	**50.22**	**47.61**	28.67	34.69	49.33	73.86
Recovery_10 [%]	34.66	81.49	**46.23**	**83.05**	44.54	51.40	**89.51**	**96.67**	40.46	89.62	70.29	93.73
Recovery_15 [%]	66.56	95.16	**42.06**	**94.81**	47.41	89.8	**83.57**	**100.22**	64.91	98.25	66.71	95.69
t_50 [min]	7.21	8.26	**2.54**	**7.49**	NaN	9.89	**4.72**	**7.25**	11.36	7.77	4.29	2.63
t_95 [min]	NaN	13.17	**NaN**	**13.4**	NaN	NaN	**NaN**	**8.59**	NaN	11.43	NaN	7.77

BMT56F

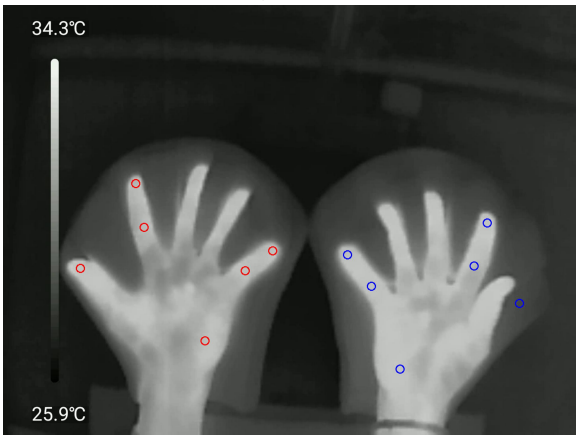
Figures C.1 and C.2 show example frames and ROI temperature over time for video BMT56F. The performance report outcomes are as follows: all before-cooling coordinates are tracked by MediaPipe, whereas the model fails in 12% of the left-hand landmarks and 2% of the right-hand landmarks after-cooling. In 28% of the before-cooling frames, the temperature cannot be determined due to the inability to extract the minimum temperature value. For the after-cooling stage, the minimum and maximum temperatures cannot be read in 9% and 3%, respectively, resulting in 11% frames with undetermined ROI temperatures. Ultimately, the total time to run the model was 31 minutes.



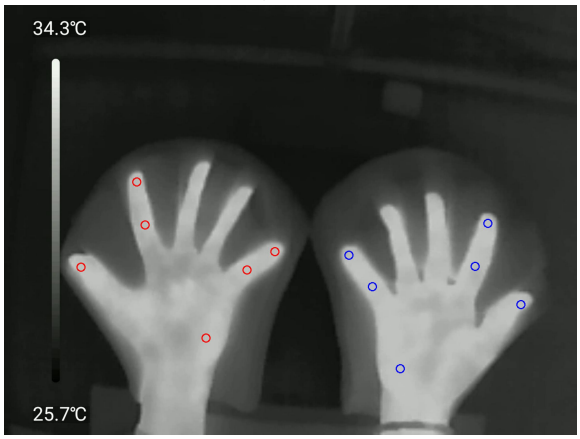
(a) Before cooling: frame 273, t = 23 sec



(b) After cooling: frame 976, t = 82 sec



(c) After cooling: frame 8053, t = 675 sec



(d) After cooling: frame 10493, t = 880 sec

Figure C.1: Annotated ROIs in frames of video BMT56F: red for the left hand, blue for the right hand

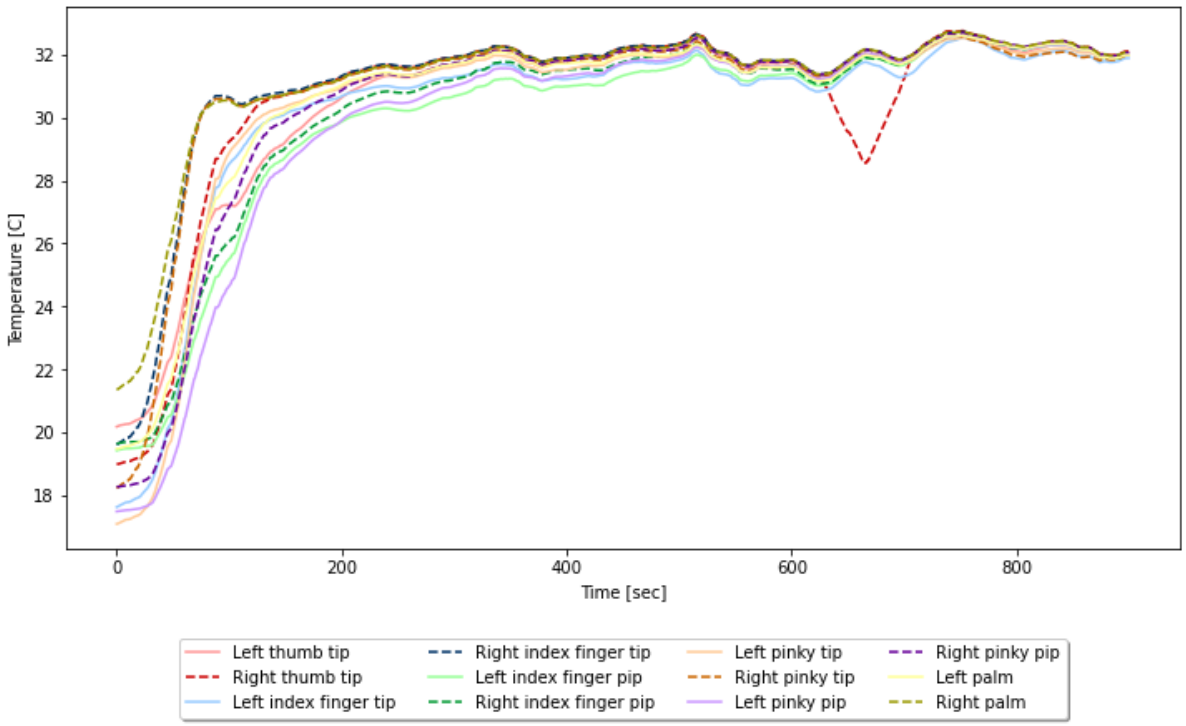
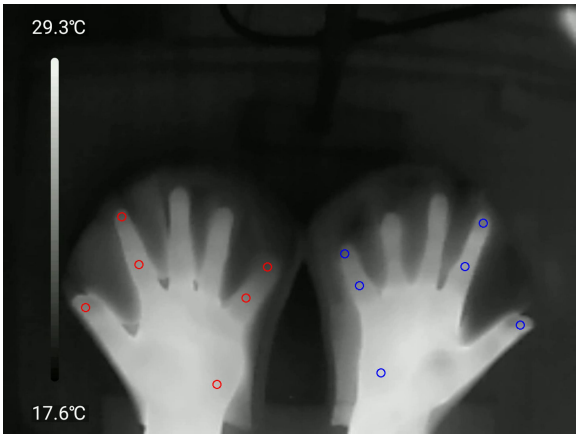
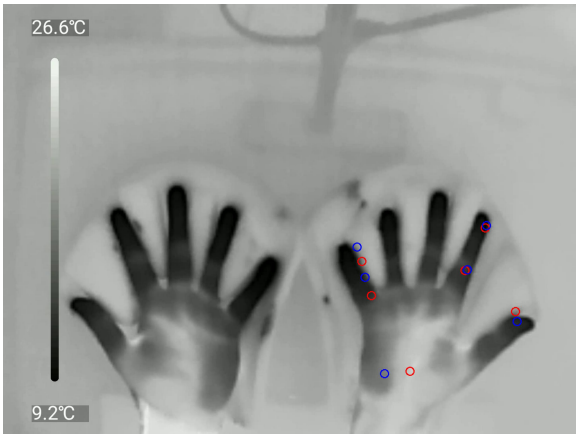


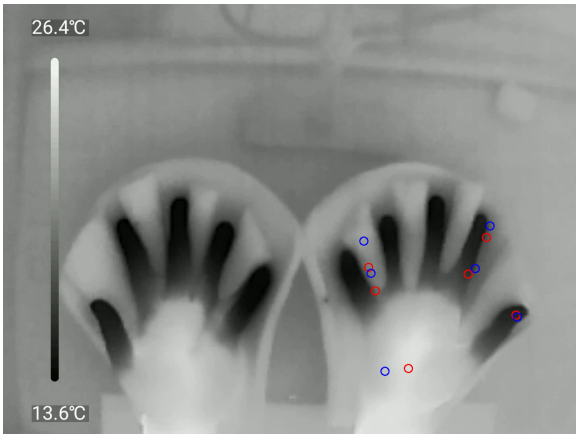
Figure C.2: Temperature development over time for twelve ROIs of video BMT56F



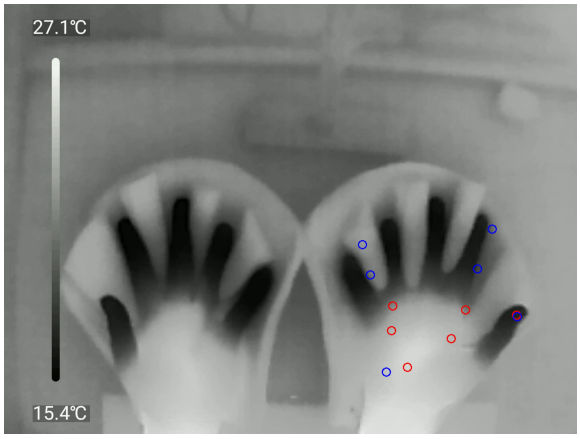
(a) Before cooling: frame 58, t = 5 sec



(b) After cooling: frame 266, t = 21 sec



(c) After cooling: frame 5869, t = 427 sec



(d) After cooling: frame 11739, t = 945 sec

Figure C.3: Annotated ROIs in frames of video SBK22M: red for the left hand, blue for the right hand

SBK22M

Four saved frames with the drawn circular ROI regions in blue for the right hand and red for the left hand, are provided in Figure C.3. The resulting temperature curves after CPT can be observed as well (Figure C.4). In the before-cooling stage, no problems related to landmark tracking or temperature extraction are encountered. After cooling, 26% and 78% of the left and right-hand landmark coordinates were not able to be found by MediaPipe. In 26% of the frames, the temperature could not be determined, caused by 2% and 26% of missing minimum and maximum temperature values, respectively. It took the model 57 minutes to complete for video SBK22M.

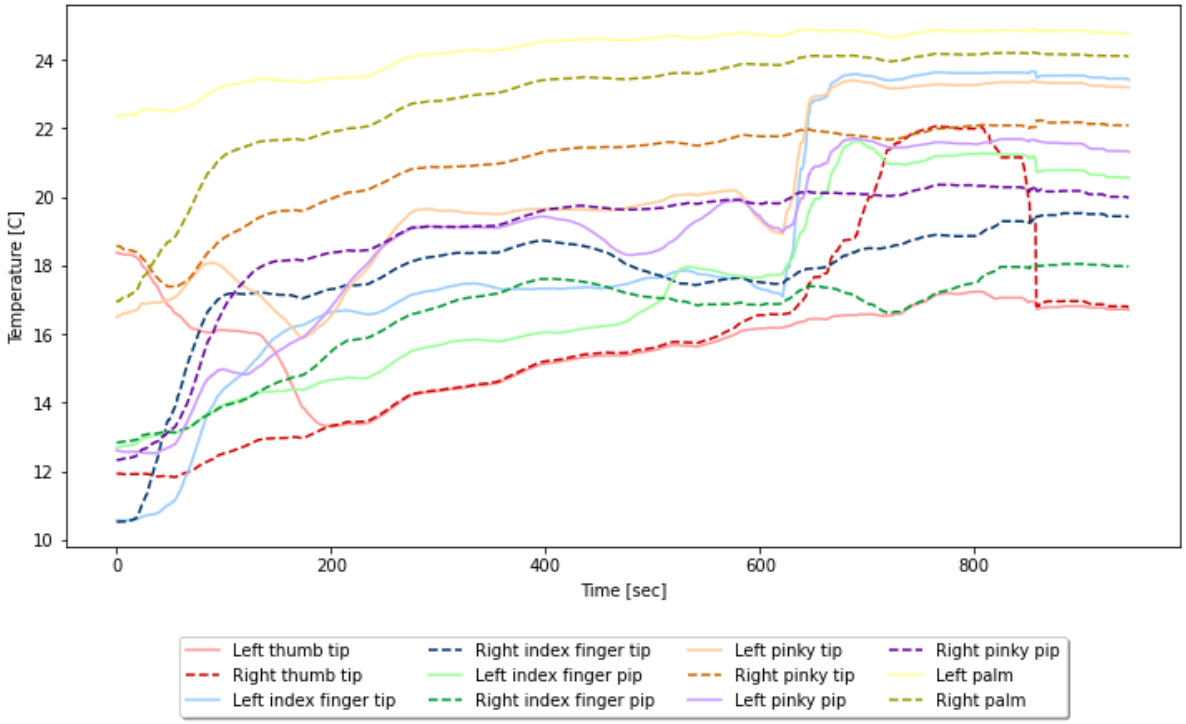


Figure C.4: Temperature development over time for twelve ROIs of video SBK22M

Repeatability

Table C.5: Intraclass correlation coefficients with p-values for the variables of interest of n = 7 videos, recorded using the classical experiment setup

Variable	ICC	p-value
Temp_Base	0.999	4.26×10^{-9}
Temp_CPT	0.995	1.78×10^{-7}
Temp_5	0.998	8.20×10^{-9}
Temp_10	1.000	5.91×10^{-17}
Temp_15	1.000	4.53×10^{-15}
Temp_Drop	0.997	5.12×10^{-7}
Recovery_5	0.994	2.53×10^{-6}
Recovery_10	0.994	3.03×10^{-6}
Recovery_15	0.995	1.82×10^{-6}
t_50	1.000	7.78×10^{-10}
t_95	1.000	2.15×10^{-9}

Visual examination

Table C.6: Percentage of before-cooling frames with correctly annotated ROIs, based on the visual examination of n = 15 videos with a regular experiment setup

	Left thumb tip	Left index finger tip	Left index finger pip	Left pinky tip	Left pinky pip	Left palm	Right thumb tip	Right index finger tip	Right index finger pip	Right pinky tip	Right pinky pip	Right palm	Mean ± Std. Deviation
AP50F	0	0	0	0	0	0	100	100	100	100	100	100	50 ± 52
BMT56F	100	100	100	100	100	100	100	100	100	100	100	100	100 + 0
IKS14F	100	100	100	0	0	100	0	0	0	0	0	0	33 + 49
KBS34M	100	100	100	100	100	100	80	100	100	100	100	100	98 ± 6
KP17M	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
MS21F	20	80	80	60	40	40	100	100	100	100	100	100	77 ± 29
P-H-EB29M2	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
P-H-R22M2	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
P-H-RA29M2	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
P-L-SB22M1	0	0	0	0	0	0	100	100	100	100	100	100	50 ± 52
P-L-SB22M2	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
P-L-SUB22M3	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
P-L-SUN29M3	80	80	80	80	100	100	100	100	100	100	100	100	93 ± 10
SBK22M	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
SG30M	100	100	100	100	100	100	100	100	100	100	100	100	100 ± 0
Mean ± Std. Deviation	80 ± 39	84 ± 35	84 ± 35	76 ± 41	76 ± 42	83 ± 37	92 ± 26	93 ± 26	93 ± 26	93 ± 26	93 ± 26	93 ± 26	87 ± 7

Effect of coordinate filtering

Table C.7: Euclidean distances between the unfiltered and filtered landmark coordinates after cooling

	Mean Euclidean distance	Maximum Euclidean distance	Euclidean distance std. deviation
AP50F	0.05	0.75	0.14
BMT56F	0.02	0.78	0.10
IKS14F	0.11	0.75	0.18
KBS34M	0.01	0.39	0.02
KP17M	0.04	0.76	0.13
MS21F	0.02	0.74	0.07
P-H-EB29M2	0.01	0.80	0.03
P-H-R22M2	0.00	0.65	0.01
P-H-RA29M2	0.07	0.78	0.18
P-L-SB22M1	0.05	0.78	0.15
P-L-SB22M2	0.01	0.63	0.03
P-L-SUB22M3	0.01	0.75	0.04
P-L-SUN29M3	0.01	0.74	0.02
SBK22M	0.06	0.76	0.15
SG30M	0.03	0.74	0.09
Mean ± Std. Deviation	0.03 ± 0.03	0.72 ± 0.10	0.09 ± 0.06