Imperial College London

Non-contact measures to monitor hand movement of people with rheumatoid arthritis using a monocular RGB camera

Letizia Gionfrida

March 2022

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

Faculty of Engineering
Department of Bioengineering
Imperial College London

Acknowledgements

First, I would like to express my sincere gratitude to my son, Adrian Cavaliere, who gave me the strength to complete this intense journey. I also would like to express my appreciation to my dear husband, Iosepho Cavaliere, which is always lifting me and supporting me as a stable lighthouse in the storm. Thank you for believing in us, believing in me, and believing in our future together.

I also would like to thank my supervisor, Professor Anil Anthony Bharath, for recognising something in me and encouraging me to evolve and grow personally and professionally. He coached me with trustworthy encouragement and sustained me at every step of the way. He is a mentor that I hope I will have the luxury to maintain for the rest of my life.

I also would like to thank my co-supervisor, Dr Angela E. Kedgley, that enabled me to gradually develop as a professional individual, teaching me meticulousness diligence while challenging me and inspiring me as a role model throughout this journey.

I also would like to thank Dr James Galloway, consultant rheumatologist at King's College Hospital, that contributed to this study. His passion for clinical practice is a trustworthy source of inspiration.

Finally comes my family. I want to take my father, Giovanni Gionfrida, for saving me and allowing me to have an education and a future, for being both soft and tender and always present for me. I also would like to thank my mother, Filomena Bernardo, for consistently believing in me and giving me more than one opportunity to prove that I was worth all the efforts and sacrifices they made since they saved my life. I also would like to thank my sweet auntie, together with my husband's family, and finally, my friends, who are like family to me.

Dedication

To Adrian, I wish you a life full of strength, love, confidence, compassion, and love.

To Iosepho, thanks for your enthusiasm, happiness, and light that radiates and completes my life.

Disclaimer

The author declares that this thesis is her original work. None of the material present in this thesis was previously submitted for a degree to any awarding institution.

Copyright Declaration

Abstract

Hand movements play an essential role in a person's ability to interact with the environment. In hand biomechanics, the range of joint motion is a crucial metric to quantify changes due to degenerative pathologies, such as rheumatoid arthritis (RA). RA is a chronic condition where the immune system mistakenly attacks the joints, particularly those in the hands. Optoelectronic motion capture systems are gold-standard tools to quantify changes but are challenging to adopt outside laboratory settings. Deep learning executed on standard video data can capture RA participants in their natural environments, potentially supporting objectivity in remote consultation.

The three main research aims in this thesis were 1) to assess the extent to which current deep learning architectures, which have been validated for quantifying motion of other body segments, can be applied to hand kinematics using monocular RGB cameras, 2) to localise where in videos the hand motions of interest are to be found, 3) to assess the validity of 1) and 2) to determine disease status in RA.

First, hand kinematics for twelve healthy participants, captured with OpenPose were benchmarked against those captured using an optoelectronic system, showing acceptable instrument errors below 10°. Then, a gesture classifier was tested to segment video recordings of twenty-two healthy participants, achieving an accuracy of 93.5%. Finally, OpenPose and the classifier were applied to videos of RA participants performing hand exercises to determine disease status. The inferred disease activity exhibited agreement with the in-person ground truth in nine out of ten instances, outperforming virtual consultations, which agreed only six times out of ten.

These results demonstrate that this approach is more effective than estimated disease activity performed by human experts during video consultations. The end goal sets the foundation for a tool that RA participants can use to observe their disease activity from their home.

Table of contents

Table of figures

Table of tables

Table of abbreviations

ACR	American College of Rheumatology
AIM	Automatic Identification of Markers
ASSH	American Society for Surgery of the Hand
CNN	Convolutional Neural Network
CRP	C-reactive protein
DAS-28	Disease Activity Score (28 Joints)
DIP	Distal Interphalangeal
DMARD	Disease-Modifying Anti-Rheumatic Drug
DoF	Degrees of Freedom
ESR	Erythrocyte Sedimentation Rate
EULAR	European League Against Rheumatism
GAN	Generative Adversarial Network
GH	Global Health Assessment
HAWK	Hand & Wrist Kinematics
IR	Infrared
LED	Light Emitting Diode
LoA	Limits of Agreement
LSTM	Long Short-Term Memory
MCP	Metacarpophalangeal
MSK	Musculoskeletal
NICE	National Institute for Health and Clinical Excellence
NRAS	National Rheumatoid Arthritis Society
PIP	Proximal Interphalangeal
QTM	Qualisys Track Manager
RA	Rheumatoid Arthritis
ReLU	Rectified Linear Unit
RGB	Red, Green Blue color model
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
ROM	Range of Motion

SD	Standard Deviation
SMA	Simple Moving Average
SJC	Swollen Joint Count
TAF	Total Active Flexion
TAM	Total Active Movement
TJC	Tender Joint Count

Chapter 1 Introduction

1.1 Background and motivation

The human hand is an essential structure for achieving various tasks of daily living (1). Its movements play a fundamental role in an individual's ability to interact with their surroundings (2). Therefore, quantifying hand movements is essential to understand hand movement disorders (3). The examination of the kinematics of the human hand can help estimate changes due to disease or treatment.

Vast is the literature to describe and measure human hand motion. Goniometers are often adopted in clinical practice. However, the reliability of goniometric measures has been discovered to be variable (4) and often unreliable (5–7). Glove (8,9) and motion sensing devices, e.g., Kinect (Xbox 360 Sensor Microsoft) cameras (10), have been tested to capture hand movements, but have the drawback of intensive manual-postprocessing, which limits the generalizability of these approaches. Traditional optoelectronic marker-based motion capture systems are considered gold-standard measurements (11) to calculate hand movement. They necessitate the attachment of markers to the participants' skin (12), positioned in accordance with the anatomy of the human hand, which serves as reference points (landmarks). These tracking systems produce accurate results when tracking the small joints of the hand (13); however, the practicability of embracing these methods outside the laboratory is restricted by their capture volume, camera resolution, the impracticality of the presence of markers during many activities, and financial factors (10).

In the past few years, markerless hand tracking methodologies to capture and quantify hand kinematics have seen significant developments (14–16). Hand pose estimation from markerless visible inputs has been a long-standing active research field. Its attractiveness is progressively increasing due to the introduction of inexpensive red green blue (RGB) cameras and the capabilities of deep learning methodologies (16,17) that have permitted better precision and robustness. These advances in video-based pose estimation have allowed automatic inference of two-dimensional and three-dimensional centres of the joints directly from video recordings. Furthermore, the accuracy of selected markerless, open-source software for pose estimation has been compared against gold-standard measurements (18–20) to enable the tracking of human locomotion adoption outside laboratory settings. However, relevant deep learning architectures and models have only been validated for the lower limb, and a critical demand to compare these approaches for the human hand remains.

Adopting markerless pose estimation models to capture and identify human limbs has generated the need for action recognition models to automatically identify where actions occur in extended video sequences and reduce manual labelling (21,22). Comprehending human movements in video recordings has a pivotal role in numerous applications (23), including classification, segmentation and content-based annotation approaches for feature extraction. However, the understanding of human hand activities within a continuous video sequence remains a complicated undertaking due to the extensive variability of images on a frame-by-frame basis (24) and the uncertain boundaries of hand gestures (25), with several studies suggesting that both temporal segmentation and classification ought to be performed in parallel with continuous gesture recognition (24–27).

A few studies have presented a workflow for translating these deep architectures and models to clinical populations to infer movement parameters and benchmark the accuracy of these models against clinical workflows (20,28–30). The ability to obtain motion metrics remotely from users in their natural environment becomes extremely attractive in clinical research to investigate effective interventions that could decrease the burden on participants with movement impairments. Capturing mobility endpoints remotely is beneficial for assisting chronic degenerative pathologies, such as rheumatoid arthritis (RA) (31).

RA is an autoimmune disease (32) in which the body's immune system attacks the joints, making them swollen, stiff, and painful. Unmonitored RA leads to synovial joint damage that results in progressively swollen joints, causing a reduced range of motion (ROM)

(33). The condition's complications can lead to permanent deformity and destruction of the joints, particularly for the interphalangeal joints of the hand (34), where RA often starts. The National Institute for Health and Care Excellence (NICE) guidelines for RA recommend occupational therapy and hand exercise interventions based on disease activity levels to minimise the loss in mobility (35). These guidelines propose intervention strategies where the best clinical outcomes are achieved through face-to-face monthly monitoring to determine if the disease activity is active or in remission. Based on disease progression, the treatment is amended and optimised. Frequent monitoring aims to identify possible flares (36), typical of the condition, and amend, interrupt or improve possible interventions.

The American College of Rheumatology/ European League Against Rheumatism (ACR/EULAR) recommendations for RA (37) show that hand ROM is a good indicator of generalised disease activity and tailored hand exercises programmes can support recovery interventions (e.g., recovering loss in mobility). Strategies that have been implemented to capture hand movements for this population include instrumented gloves (38,39), motion-sensing devices (40), and optoelectronic motion tracking systems (41). However, there are problems associated with embracing these approaches in clinical practice. Gloves may limit the joints' ROM. Motion-sensing devices and optoelectronic motion tracking systems are applicable in only limited scenarios and are challenging to deploy remotely in daily activities. Evidence suggests there is a vital clinical need for a monitoring technology that can remotely measure clinical endpoints of disease progression for RA patients, an approach less user-dependent and that needs less assessor input (31,42,43).

The need for remote strategies that can replace face-to-face assessments is emphasised by the increased incidence of RA, causing a growing demand for rheumatologists to monitor this condition, despite a workforce shortage in this field (44). In 2009 the National Audit Office revealed that only 10% of RA patients received adequate follow-ups, consisting of monthly in-person assessments to determine disease activity. This issue has been aggravated by the COVID-19 pandemic that has dramatically affected rheumatology workflows, including services and patient interactions (45). In-person consultations and face-to-face in clinics, once routine, have been replaced by virtual consultations that, although generally accepted (46,47), have struggled to capture objectively clinical endpoints of disease progression in RA.

To address these needs, a remote monitoring tool that uses a single monocular RGB camera and the above-mentioned deep-learning-based algorithms could be leveraged to capture and assess the ROM of the small hand joints in RA patients. This approach could support current remote monitoring procedures to acquire objective clinical endpoints in RA disease progression, improve ongoing virtual consultations, and facilitate more frequent monitoring from ubiquitous technologies. More regular, accurate, and systematic monitoring of disease activity status could even assist the design of more tailored interventions (48) for this population.

1.2 Research aims and objectives

This thesis aims to investigate the use of monocular RGB cameras to estimate joint ROM and assess disease activity in patients with RA.

The work presented is divided into three parts. The first part evaluates the accuracy of a markerless tracking system against the gold-standard marker-based model. Here, the hand ROM of healthy volunteers captured using the two tracking modalities is compared. Furthermore, as part of this comparative evaluation, several filtering techniques and different image enhancement visualisations are examined to consider extreme cases in which the selected markerless tracking did not perform well.

The second part of this thesis describes the implementation of a novel hand gesture recognition model. This model aims to classify and segment continuous video recordings to identify only the subsets of the videos where relevant hand gestures occur. On the extracted subsegments, the above-validated markerless tracking was then executed.

In the third and last part of the thesis, the previous models, implemented and validated on healthy volunteers, are extended to RA patients. The disease activity obtained a priori in the clinic (also called *ground truth*) is compared against the disease activity; i) visually estimated by a clinician over a remote virtual consultation, ii) inferred by the markerless models based on hand joint ROM.

The specific objectives were to:

1. compare and validate a state-of-the-art markerless hand pose estimation method in capturing finger joint ROM against the gold-standard marker-based optoelectronic motion tracking.
2. develop a novel gesture recognition model to automatically classify and segment extended video sequences to assist the markerless ROM detector.
3. expand these models to patients with RA, evaluating the capability of these models against current virtual consultations to assess RA disease activity.

These goals are investigated in specific chapters, as follows.

1.3 Thesis outline and chapter summary

Chapter 2 offers an overview of optical-based approaches utilized for tracking, measuring, and configuring hand ROM. It moves from goniometric assessments, glove-sensing devices, and motion-sensing devices to active and passive marker-based and markerless hand pose estimation methods. Finally, it also presents an overview of active and passive hand gesture recognition models and how they have been used to segment and classify hand gesture signatures from continuous video recordings.

Chapter 3 gives an overview of the epidemiology and costs of RA. This work discusses how RA is assessed in the clinic, its limitations, and how ROM has been measured for this population over the last decades. A comprehensive review is conducted to evaluate the most well-known published studies on tracking hand ROM in RA. Finally, this chapter outlines alternative remote monitoring tools used to capture clinical endpoints for this population and how they have been used to support virtual consultations.

The aim of Chapter 4 is to validate a methodology that enables markerless hand pose estimation for healthy volunteers, comparing a markerless approach to a marker-based system. Here, hand ROM is acquired from a commercially available marker-based motion capture system synchronized with an RGB camera. Following data labelling and filtering, the measured phalangeal ROMs extracted using a marker-based technology are compared with phalangeal ROMs obtained with the markerless approach. The technique analyzed in this chapter presents the error in degrees, suggesting the usability of such a markerless technique applied in clinical practice.

Chapter 5 introduces a novel hand gesture recognition model, developed to identify a set of defined hand gesture signatures when capturing colour video frames, with the goal of optimizing markerless hand tracking techniques. Here, an action detection classifier that looks at both appearance and spatiotemporal parameters of consecutive frames is illustrated. To leverage the need for large-scale dataset to train a deep-learning architecture, the implemented network uses an available open-source dataset. Furthermore, it uses a technique known as transfer learning, to fine-tune the model on the hand gestures of relevance in the clinical context of RA.

In Chapter 6, the models developed in Chapters 4 and 5 are extended to deliver a remote monitoring proof-of-concept that captures interphalangeal joint ROM and estimates disease activity. Chapter 6 presents the outcome measures over this cross-sectional investigation. The clinical data, including ground truth disease activity captured in the clinic, visual examination performed over a virtual consultation and inferred joint ROM, are collected, and assessed. This chapter compares the results of the assessment conducted in clinic when all the components were examined, including blood tests, pain measures and joint assessment, against the pipeline implemented in Chapter 4 and Chapter 5, and a virtual consultation where a rheumatologist assessed the disease activity based on visual examination.

Finally, Chapter 7 summarises the main findings from each chapter and discusses the implications of these findings. This includes providing indications on the impact these results may have on the management of RA. This chapter also summarises the main strengths and limitations and the contribution to the field. The methods adopted to conduct the analyses are also discussed, suggesting possible improvements and technical constraints. Recommendations for future research suggests the need for a longitudinal clinical investigation. The latter would look at individual components of disease activity score in RA.

Chapter 2 Optical-based measurements for quantifying hand kinematics

2.1 Introduction

The hand is the most intricate anatomical structure in the human body, and its mobility is essential to enable interaction with the surrounding environment (49). A functional human hand can deliver a broad range of motions (ROMs) to perform different tasks (50). Decreased hand mobility may happen for diverse reasons. Common causes include finger injuries (51) and ageing (52). However, hand motion can be compromised also by chronic pathologies that affect the nervous system, like Parkinson's disease (53), or the immune system, like rheumatoid arthritis (RA) (54). To assess human hand impairments, clinicians often use visual examinations or clinical grading scores. However, these assessments rely on the evaluator's experience, which affects cross-comparison reliability of diagnosis (54). An unbiased quantification is desirable to preserve and possibly enhance, clinical decision-making.

To provide objective quantification of human hand movements, an individual's hand can be expressed as a range of rigid multibody mechanisms (55), made of a collection of segments linked by joints. These articulations connecting the segments have single or multiple degrees of freedom (DoF) (56). The human hand has 27 DoF (57), involving a total of four bones for each finger, including metacarpals, proximal phalanges, middle phalanges, and distal phalanges (50) (Figure 1).

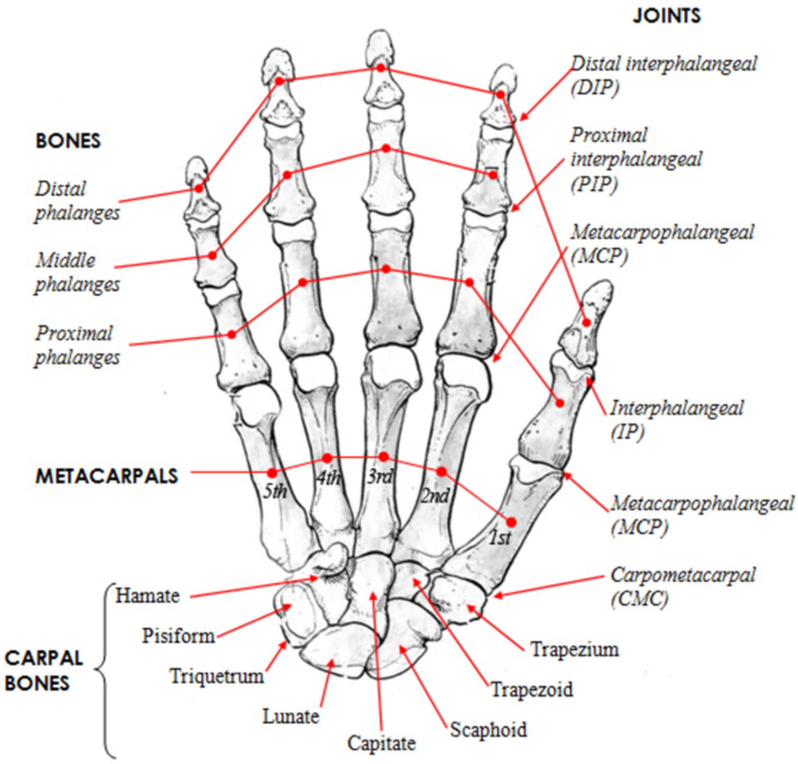


Figure 1: Human hand skeletal structure illustrating the phalanges bones and joints. Reproduced with permission from (58).

The articulations between the metacarpals and proximal phalanges are called the metacarpophalangeal (MCP) joints. The articulation between the proximal and middle phalanges is the proximal interphalangeal (PIP) joint. The joint between the middle and the distal phalanges is called the distal interphalangeal (DIP) joint (59). Thumbs do not possess a middle phalanx; hence, they have an MCP and a single interphalangeal (IP) joint (50). Of the 27 DoF, the 2^{nd}–5^{th} fingers each have four DoFs, including two DoFs for the MCP joint and one DoF for each of the PIP and the DIP joints, while the thumb has five DoFs, with six DoFs for the wrist, three for rotations and three for translations(57). Once the DoFs are recognized, inverse kinematics may be applied to locate the individual articulations and reconstruct the full kinematic chain (55).

A large variety of instrumented tools have been used to objectively quantify human hand kinematics (57). Amongst the many approaches utilized in clinics, goniometers (Figure 2) are broadly accepted for statically measuring hand movements (60). Depending on the joint under examination, several types of goniometers (5,61,62) can be chosen, varying in size and shape. Goniometers offer a low-priced and transportable solution to quantifying ROM (5,7,62). Nonetheless, they can be affected by lack of reproducibility (4); thus, new approaches to measurement have been explored.

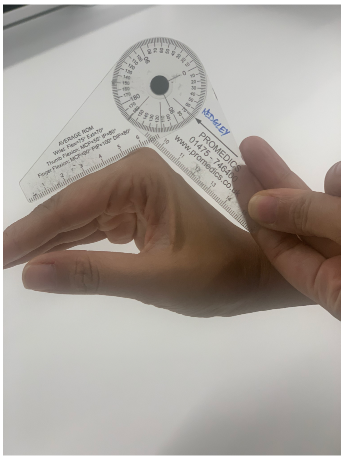


Figure 2: Illustration of a short arm universal goniometer for finger joint angle used to measure the metacarpophalangeal joint angle of the index finger.

Alternative methods to capture hand movements include acoustic, mechanical, magnetic, and optical systems (63). Visual-based methods aim to provide minimal obstruction while working with single or multiple cameras to locate human joints (64). These tracking devices have been categorised based upon their working principles, dividing these methods into marker-based and markerless (65). Marker-based systems can operate using two working principles: active tracking, by attaching a light source to the user's skin, and passive tracking, by affixing reflective markers to the user's skin. Both methodologies use infrared (IR) cameras (55). These optoelectronic systems, properly utilized, can produce accurate results and are frequently employed as the gold standard measurement to quantify movements (66). However, the practicability of using these methods within a home surrounding is restricted. The adoption of these practices continues to be confined to laboratory settings for manifold reasons, including the physical space and economic limitations (10).

The field of optical markerless hand tracking has seen significant advancements with novel deep-learning-based methodologies published from 2014 (67). This is due to the ubiquity of low-priced RGB cameras and the introduction of deep convolutional neural networks (CNNs), which have facilitated heightened accuracy levels and robustness (68). CNNs (69) are a class of neural networks, most traditionally used to analyse image frames. Using CNNs, open-source libraries, such as OpenPose (17) and RGBNet (16) have been implemented, exhibiting great potential to recognize hand joint centres from two dimensional color frames. These markerless approaches can provide estimates of the declination of joint angles, producing accurate results (62,70,71). However, even if the accuracy of these open-source tools continuous to grow, their adoption in the field of hand biomechanics has not yet occurred (72), as the finessing of such algorithms is usually outside the field of traditional hand biomechanics.

Notwithstanding the potential of these markerless solutions, some hesitancy in the biomechanics community is present (72). To ease this translation process, a few procedures have recently emerged to embrace these markerless estimation procedures in lower limb biomechanics, as the community requires validated accuracy to adopt such solutions (18,73). For instance, using more than one RGB camera to reduce the occlusion introduced by other body segments, exploiting a direct linear transformation (18) and triangulation techniques (73). These novel methodologies have benchmarked the accuracy three-dimensional kinematics captured using these markerless systems (e.g., OpenPose) against three-dimensional kinematics captured using gold-standard marker-based passive optical capture technologies (18). However, introducing more cameras would limit the reproducibility of a study in home settings.

Therefore, recent investigations have compared the two-dimensional kinematics inferred by adopting these techniques against more traditional three-dimensional motion tracking systems (19,74). The accuracy and validity of these approaches for the lower limb have encouraged the adoption of these algorithms to quantify gait kinematics also on impaired participants (20,28). These solutions have illustrated that markerless motion capture is a practical instrument that can augment current research opportunities, rather than acting as an outright substitute for conventional laboratory-based strategies. However, these studies have only validated markerless tracking technologies for the lower limb, and the question of whether they could be embraced for hand kinematics remains open.

2.2 Optical systems to quantify joint range of motion

In the late 1950s, Noer and Pratt (75) introduced a measurement type of protractor, called a goniometer, to estimate the declination of phalanges joint angles. The preliminary version for measuring the ROM, issued by the American Academy of Orthopedic Surgeons (AAOS) in 1965, recommended that estimating joint angles was influenced by more considerable inter-reader variability than goniometric assessments (76). By then, the goniometer had become omnipresent in clinical practice (76). In modern practice, health care professionals frequently use goniometers to estimate joint ROM (60). The goniometer (Figure 2) is utilized to estimate and assess both a the angle of a specific joint pose and the entire ROM (77). The assessor collects the measures by locating the goniometer along with the articulations estimated (78). These instruments support clinical decision-making and assist health care professionals in monitoring their patients to maximize and enhance the outcomes (79). One type of goniometry instrumentation extensively used in hand therapy and occupational therapy is called the universal goniometer (Figure 2) (80).

Research to assess the reliability of universal goniometers has described low inter-reader and intra-reader reliability, with high variability (4,81,82). In Macionis' study (7), the errors of measurements of universal goniometers ranged from 2.4° to 4.9°. Other studies on the unsteadiness of universal goniometers stated variation of 7°–9° amongst therapists when measuring joint angles (83,84), leading to a 27° variance in the phalanges articulations. In Somers et al.'s study (85), it was highlighted that a large part of the reliability variation in universal goniometers is due to the influence of experience in goniometric assessments.

One of the difficulties in modern practice is that these tools demand physical contact with the finger to obtain the most reliable precision (50). However, tissue injuries and dermatological conditions can generate challenges with practical use because of the danger of contamination, bandages, or discomfort. With the development of new technologies, new goniometer models have been gradually introduced and improved to assist clinicians (6,86,87).

A substitute to conventional goniometric estimations is the application of visual tracking methods. Several approaches have been applied to recognise the hand's gesture and shape. The most commonly trusted arrangement for hand motion capture (88) employs optical sensing technologies.

Optical systems for use in movement capture include those using IR light-emitting diodes (LED), depth cameras, and RGB cameras; single or multiple cameras (64) may also be used to sense position. There are two commonly used techniques of optical motion capture (8). The first methodology applies markers to the human body, while the second does not (10). In both cases, a set of two or more cameras is located around the limb of the movement to be analysed. Software then associates the many perspectives and uses camera intrinsic and extrinsic parameters (e.g., focal length) to estimate three-dimensional coordinates for the objective of interest.

2.2.1 Marker-based

In marker-based optical systems, markers are attached to the segments of interest (64). Two types of marker-based optoelectronic systems can be used, including active, where markers are IR LEDs (63), and passive modes (66). Active systems utilize markers that emit light (66) (Figure 3). These brightened markers function as the main signal source and are often deployed using IR LEDs. The application software prevents the swapping of markers with one another and identifies them also after occlusion.

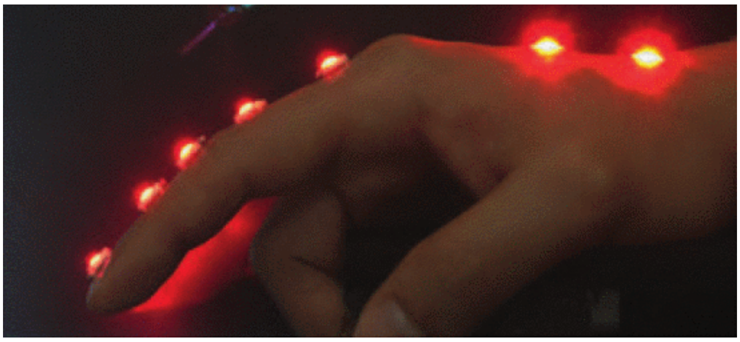


Figure 3: Example of an active optical motion-capture set-up. Reproduced with permission from (89).

Active markers provide measurements that are considered more reproducible than standard goniometry assessments (66). However, they come with several limitations.

They may require cables or batteries (90) because each LED indicator needs to be powered or wirelessly recorded, in which case users have to put on wearable energy packs or be tethered to the motion capture system via a system of wires. This results in limited freedom of movement, particularly when tracking small motions, as in the case of the joints between the phalanges. Moreover, given that each lighted marker demands a distinguishable frequency rate to be recognised, the maximum limit of the frame rate may decrease if various LED frequencies have to be multiplexed (64).

Passive marker-based tracking modalities are more commonly used to quantify hand kinematics (91,92). These systems track kinematics with passive retro-reflective markers placed on pre-defined landmarks (93), enabling three-dimensional motion tracking and analysis (94). Data are acquired with multiple cameras, usually placed such that they are equally spaced around the capture volume (Figure 4A). The high-resolution cameras have active IR LEDs around their lenses, matched to a strobe light controller to recognize reflective markers across time (95). As the human body performs a motion in the capture volume, reflective markers return the incoming IR light in the camera lens, triggering a photosensitive plate that enables the recordings (72).

(A) (B)

Figure 4: (A) The optoelectronic motion capture (Qualisys AB, Gothenburg, Sweden) system being used to collect the hand kinematics adopting the passive marker-based setup. (B) Example of a type of marker set used in kinematic studies of the hand.

Human upper body motion is highly variable and complex (96), and such variability and complexity makes it difficult to standardize hand movement (97). Thus, it is important to position the markers in strategic positions (55), with experimental protocols often presenting different marker configurations developed specifically to assess hand kinematics (98) (Figure 4B).

Anatomical marker sets use bony anatomical landmarks as the attachment sites of single markers (60,99,100). Anatomical landmarks are characterized by a bony prominence and require marker placement via palpation procedures (101). As palpation errors of up to a few millimetres have been shown to result in angular errors of several degrees, significant inter-participant differences can pose a threat to set-up reliability due to large variability (102). However, the main disadvantage is related to the fact that many landmarks are adjacent to the joints, where skin movement will be greatest, causing markers to move non-rigidly with respect to the underlying bones (103). One variant of marker-based techniques is the use of clusters of markers that are mounted on a rigid or semi-rigid plate.

For some lower limb segments, this technique has been shown to reduce artifacts related to soft tissue movement over bone, as only the markers on the plates are tracked during functional tasks (104). However, there is limited evidence on the use of marker clusters to reduce the incidence of soft tissue artifacts in the upper limbs (96).

These passive optical technologies have the advantage of being able to cover large areas (105). Another advantage is the possibility of collecting data over extended time intervals (106). This technique has the advantage of being highly accurate compared to goniometry assessment, when capturing three-dimensional motion (106). The method is non-encumbering to the patient or participant, with the only apparatus being small low mass reflective markers, compared to electro goniometers that can heavily limit motion (72). Another great advantage of this motion tracking technology is that it is not influenced by metal or electromagnetic interference (107).

However, this capture system possesses some disadvantages, including the line of sight requirement, and greater accuracy comes at a greater price, compared to goniometers (66). Most are affected by large execution time during simulation, complicated by challenging algorithmic analysis (55). The use of such systems can be considerably time-consuming because the accurate placing of markers, one-by-one, is slow. Moreover, errors can be caused by muscle deformation during a movement and skin sliding, which occurs frequently with older people (103). Given the advantages of passive optical systems and the recent improvements in computing speed and digital cameras that have reduced the above disadvantages, video-based marker systems are now considered the gold standard in hand kinematics research (66).

Once the centres of the joints under inspection have been determined, inverse kinematics are used to assess the configuration of the joints in a kinematic chain (108). There are different methods for performing inverse kinematics (109,110). To date, the most used technique implemented in commercially available optoelectronic passive motion capture systems makes use of the Forward And Backward Reaching Inverse Kinematics (FABRIK) algorithm; FABRIK (111) is an iterative algorithm developed for tracking marker locations (112).

2.2.2 Markerless

In recent years, markerless two-dimensional and three-dimensional tracking technologies have been demonstrated to be a powerful tool for accurately estimating hand movements outside laboratory settings (113). Many technologies have been introduced to quantify two-dimensional and three-dimensional phalangeal joint angles and finger position, such as digital goniometers, and commercially portable motion-sensing devices (14) and cameras (114) linked to tracking software (115). However, tiny and hardly apparent articulations, occlusions, and lighting changes can make this approach to tracking an even more challenging problem compared to marker-based assessments (114).

2.2.2.1 Motion-sensing devices

Some recently introduced software-based digital goniometers utilize accelerometers to calculate two-dimensional joint angles (116–118). These goniometers have various advantages, including availability, facility of measurement, and one-hand usability. However, there is low transferability of the technique to clinical settings due to time requirements and instrumentation (119). To calculate two-dimensional joint angles from images, an additional issue relates to the arrangements of cameras, which can cause difficulties in detecting smaller joints (120), suggesting that automated methods based on machine learning may be needed.

Portable and commercial optical measurement systems also have been introduced to accurately assess hand kinematics (14). Capture of hand movements with ubiquitous optical devices has attracted researchers' attention, particularly with the introduction of the Microsoft Kinect Sensor (Microsoft Corp., Redmond, WA, USA) (121) and the Leap Motion Controller™ (Leap Motion, San Francisco, CA, USA) (122), both illustrated in Figure 5. These commercially available gaming and user interface systems can offer non-contact and rapid solutions compared to goniometric assessments in measuring both body and finger movements (123).

(A) (B)

Figure 5: Example of two motion sensing devices. (A) Microsoft Kinect reproduced with permission from (124). (B) Leap Motion Controller™ reproduced with permission from (123).

Microsoft Kinect consists of a collection of sensors, including an RGB and depth camera (121). This tool can detect both raw depth images and a three-dimensional virtual skeleton of the body (125). Research studies have broadened the Kinect range of capabilities, originally designed to track only the larger limb segments, to expand to a three-dimensional model of the human hand. In 2013 Metcalf et al. (10) presented a Kinect-based system to detect and assess hand mobility. However, it has been reported to have an accuracy of less than 15° when capturing finer hand phalangeal motions (50), potentially because it was designed to monitor motion of the entire body (126). Other limitations include the limited distance detection depth and the extreme sensitivity to sunlight, making the device not suitable for outdoor applications (127).

The Leap Motion Controller™ (LMC) captures three-dimensional hand movements using three IR LEDs and two RGB cameras (14). The IR LEDs track the position of the palm, wrist orientation, and the five digits (122), outputting a point cloud with one point for the centre of the palm and one for each joint. The LMC™ has been applied to a large range of applications ranging from gaming to hand gesture recognition (128,129). However, drawbacks include the fact that users cannot set the sample rate and output data are frequently missing points (128).

To address these challenges, the popularity of inexpensive depth and RGB cameras, new approaches, as well as new insights, have enabled greatly improved accuracy levels and robustness for markerless hand pose estimation (130). The state-of-the-art hand tracking methodologies have seen great advancement in the field, particularly with the introduction of convolutional neural networks (CNNs) in many methods (17,131–133), and large datasets being published (134).

2.2.2.2 Machine learning approaches

CNNs are broadly utilized for automated feature extraction from imagery content. A general CNN architecture is illustrated in Figure 6.

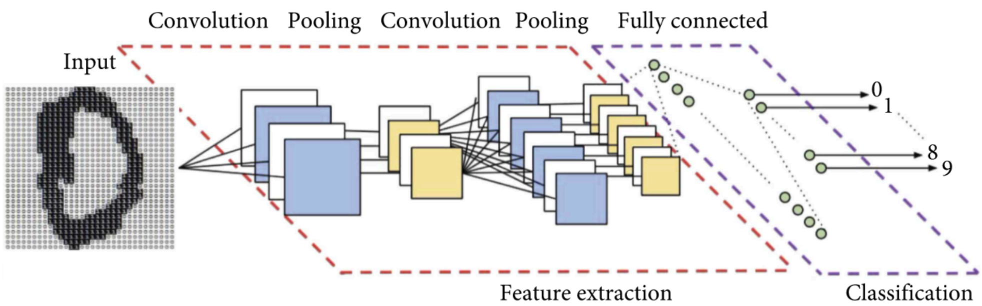


Figure 6: Example of a convolutional neural networks and its layers. Reproduced with permission from (135).

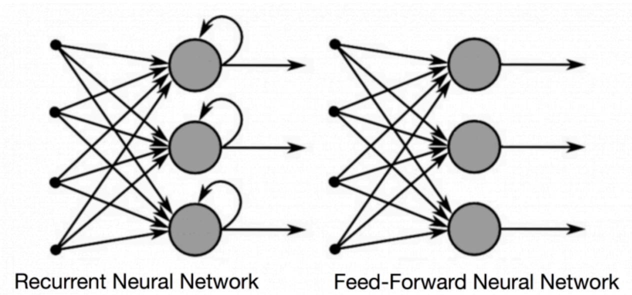
These CNN architectures embody three main steps of processing (also known as layers) that include the input layer, the feature extraction layer and output layer. The input layer takes imagery content as input and sends the data to the feature extraction layer, also known as the hidden layer. The feature layers include a pooling, and a classification layer (136). In the convolution layer, every neuron is associated with a filtering (or kernel) window that is convolved with the input. This convolution operation is comprised of individual convolution units, known as neurons, each associated with a set of parameters, known as the weights. The weights transform the input data for each associated neuron.

The output of this convolution step is a set of N images, one for each of the N neurons. Because of convolution and the use of signed weights, these new images can contain negative values. Activation functions, typically taking the form of so-called Rectified Linear Units (ReLU) are applied; one obvious effect of such units this is that they replace negative values with zero (137). In a more abstract sense, they encourage sparsity in the outputs of neuronal responses, and allow units to behave in a non-linear fashion, increasing the types of mappings performed by the network above purely linear systems.

The outputs of the convolution layers are also called feature maps. Following the convolution layers, spatial pooling layers are often included to reduce the dimensions of the feature maps and expand the context of inputs seen by deeper units. The design of many CNN architectures alternates between convolution and pooling layers, for instance, Szegedy et al. (137) presented a CNN with five convolution layers followed by one pooling layer. At the end of convolution and pooling layers, there is typically a multilayer perceptron network that performs the classification, based on the feature maps computed by the previous layers.

CNNs have been used previously for hand tracking (136,138,139). For instance, Molchanov et al. (140) and Flores et al. (141) both presented three-dimensional CNNs using depth sensors to recognize hand poses, allowing automatic extraction of diverse image features including edges, circles, lines, and texture, as well as automatic classification. These models have proven effective for feature extraction when large datasets are used. However, these architectures present some problems when looking at temporal dynamic behaviours, and therefore are combined with other architectures or layer types.

Recurrent neural networks (RNN) also have been adopted (142,143) to track dynamic motion trajectories. In an RNN, connections between nodes implicitly move along a temporal sequence. This allows the network to be responsive to time-varying information. RNNs are distinct from purely feed-forward networks, as the one described above. In purely feed-forward networks, data move only in one direction from the input layer, through the hidden layers, to the output layer. These networks have no memory of the input they receive and are not able to predict what’s coming next, considering only the current input. In RNNs, the information cycles through a loop (Figure 7). This enables RNNs to be able to recall what occurred in the training and, when deciding, to consider both the current input and the state of the inputs received previously.



(A) (B)

Figure 7: An illustration of (A) a recurrent neural network and (B) a feed-forward neural network. Reproduced with permission from (144).

Even if these architectures perform well, they can be affected by a major challenge known as vanishing gradients. A gradient is a partial derivative with respect to its inputs. It can be seen as a slope of a function, where the higher is the value, the steeper is the slope, and the faster the RNN model learns; in contrast, if the slope is zero, the RNN model stops learning. Vanishing gradients occur when the values of a gradient are too small, and the model stops learning or takes too long as a result.

The vanishing gradient problem was partially solved through the introduction of a structure called long short-term memory (LSTM), an extension of RNN. LSTMs assign weights to help RNNs let new information in, forget information, or assign importance enough to impact the output. This is because LSTMs contain information in a memory and can be seen as a gated cell. Gates determine whether to let new input in (input gate), delete the information or let it impact the output at the current timestep (output gate) (Figure 8). The gates in LSTMs take in the form of sigmoids, meaning their value ranges smoothly from zero to one. The fact that they show a smooth transition from "off" to "on" enables them to support back-propagation. Therefore, the problem of vanishing gradients is solved through LSTM because it keeps the gradients steep enough, which keeps the training relatively short and the accuracy high.

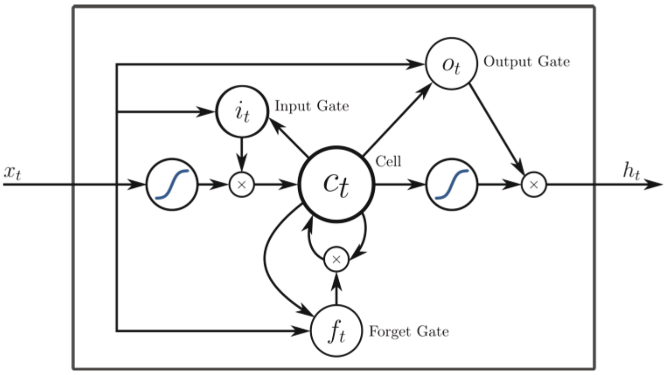


Figure 8: Long short-term memory (LSTM) structure with input ($\boldsymbol{x_t}$) and output ($\boldsymbol{h_t}$) vectors. The LSTM unit includes the input ($\boldsymbol{i_t}$), the output ($\boldsymbol{o_t}$), and the forget ($\boldsymbol{f_t}$) gates. Reproduced with permission from (145).

RNNs have demonstrated the ability to perform storage of internal states and implement complex dynamics to provide a natural framework for both recognition and prediction of temporal sequences. However, the main disadvantage of such networks is their difficulty to converge during training when the data contains complex sequential structures associated with, say, language.

Currently, the most popular hand tracking architectures tend to take advantage of the strengths of both CNNs and RNN, taking the form of hybrid CNN-LSTM architectures. CNNs and LSTMs extract the characteristics of a video, classify to which class each element belongs, and are both end-to-end trainable and real-time capable (136) (146). The most used architectures for hand tracking follow variations around those illustrated in Figure 9.

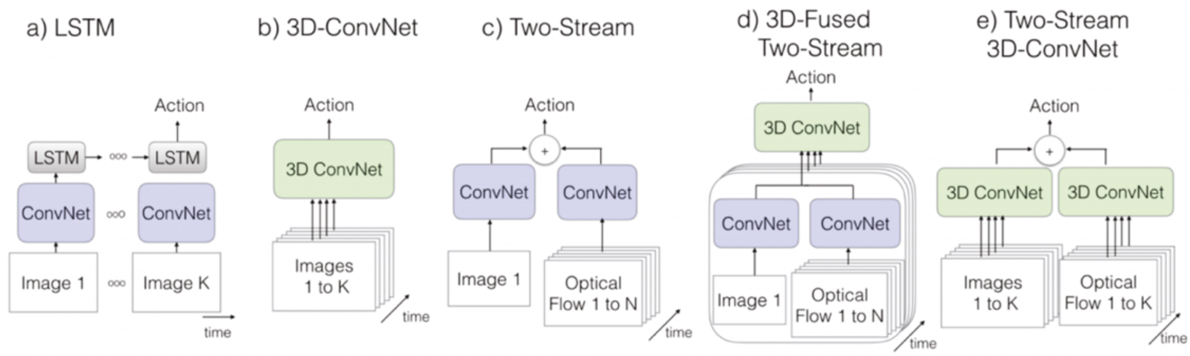


Figure 9: Architectures commonly used for hand tracking. K is the total number of frames in a video, N is a subset of neighbouring frames of the video. a) and b) use a CNN-LSTM architecture c) – e) are not real-time capable. Reproduced with permission from (147).

Using these architectures, markerless hand tracking from depth-based cameras has seen significant progress due to the low cost associated with these sensors (148). Traditionally, depth-based tracking was the main approach in human body tracking estimation (130). Several studies have suggested that these systems scale well with the size of the training dataset (15). The availability of a large-scale, accurately annotated dataset is therefore a key factor for advancing the field of pose estimation.

Markerless tracking with RGB cameras has a more extensive range of applications, considering the ubiquitous availability of these devices. However, RGB-based image networks contain less information than depth disparity maps, are more difficult to train, and require a wider set of data (149). The major challenges encountered in defining RGB-based pose estimation models include large-dimensional problems, uncontrolled environments, self-occlusions, processing speed, and rapid hand motion.

Amongst some notable recognised efforts in the field of three-dimensional markerless RGB-based hand pose estimation, there is the methodology introduced by Zimmermann et al. (16). Like other works on three-dimensional markerless pose estimation, they used a two-part pipeline. First, they detected keypoints (joint centres) in two-dimensional coordinates and then elevated the set into three-dimensions. Methodologies for moving from two-dimensional to three-dimensional space include neighbour matching (150), mixture of probabilistic principal component analysis bases (151), or direct linear transformation using multiple RGB cameras (18). Zimmermann et al. (16), first used a CNN named HandSegNet (Figure 10) to segment the hand. Then the image was cropped and resized before running a two-dimensional joint detection based on a probability density map, called PoseNet. To extract three-dimensional coordinates, a regression-based network was used. Here, the length of the distance between the joints was normalised. Ultimately, a three-dimensional matrix was applied to rotate keypoints.

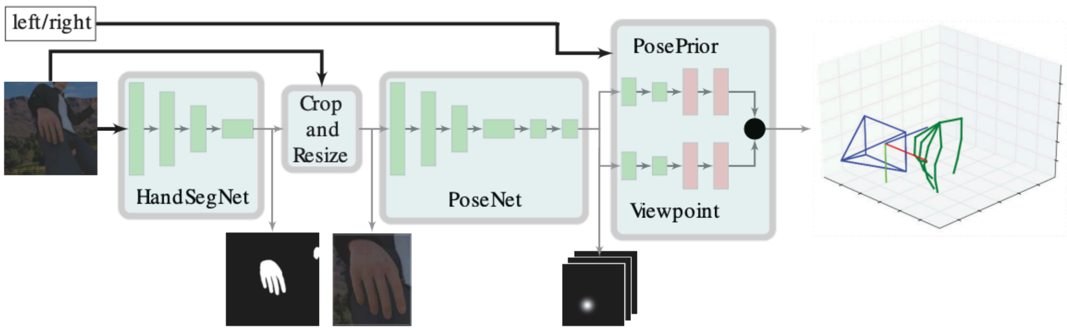


Figure 10: Proposed architecture to estimate three-dimensional hand keypoints. Reproduced with permission from (16).

While this application was proven to be effective, the latency the system shown to be excessive due to the nature of the processing pipeline (114). On top of this issue, the hand has very high-speed movement capabilities that make it difficult for many algorithms to reliably perform tracking of digits or joints across consecutive frames (149).

To address all these issues, Simon et al. (152) proposed a multicamera approach using more than 500 RGB cameras in a rounded space (Figure 11). First, they used a synthetically-generated dataset to train a hand pose model. Then, they put a volunteer in the multicamera set-up and used the previously trained hand model on each of the 500 RGB cameras while capturing motions for each view. The algorithm produced inaccurate results when the hands were occluded. To address these inaccuracies due to occlusion, they proposed a further step that they called the triangulation step. In this stage, they lifted two-dimensional coordinates to three-dimensions, knowing all the intrinsic and extrinsic parameters of the cameras. Here, they used a RANdom Sample Consensus (RANSAC) algorithm, a technique to estimate parameters by random sampling observed data. Finally, they used multiple views to project the three-dimensional views to the frame to be annotated. The authors (152) used this approach to obtain a fully annotated dataset without having to annotate data manually. Once the new dataset was obtained, they trained the original CNN three more times to obtain a model that could deal with occlusion, run fast and was able to provide good results even with rapid hand motions. However, the model only worked for two-dimensional estimation.

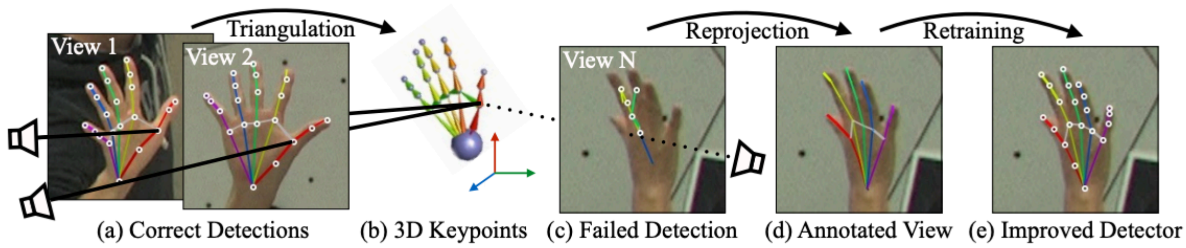


Figure 11: The process implemented in Simon et al.'s (152) work. Reproduced with permission from (152).

To test the accuracy of OpenPose (152), Nakano et al. (18) presented a study that tested its performance, the RGB-based markerless work presented by Simon et al. against a gold standard marker-based optoelectronic motion capture set-up for full body tracking (without hands), using multiple RGB cameras. They adopted a direct linear transformation (153) to elevate three-dimensional coordinates from two-dimensional keypoints detected with OpenPose. Their study illustrated an accuracy of 30 millimetres (mm) when comparing the two tracking methodologies.

To reduce the number of cameras used when testing the accuracy of OpenPose against gold standard marker-based tracking, D'Antonio et al. (73) implemented a pipeline made of two RGB cameras for lower limb motion tracking. They used a linear triangulation algorithm to convert two-dimensional coordinates obtained from OpenPose into three-dimensional coordinates. Results revealed that their approach had a root mean square error (RMSE) that was always less than 9.9°.

Following these studies, there has been a great deal of interest in validating OpenPose as a tool for markerless motion capture, particularly in gait biomechanics, using just one camera to collect data outside laboratory settings. Sakurai and Okada compared parameters of gait analysis acquired by OpenPose using just a single video camera with those obtained with a conventional system using IR cameras (154). Their study showed an error of approximately 5° between the systems observed in lower extremity joint angles. Similarly, Stenum et al. (19) matched OpenPose against three-dimensional kinematics captured with a marker-based optoelectronic motion capture. They obtained an error that was always less than 7.4° when the two methodologies were compared. Drazan et al. evaluated the performance of OpenPose in extracting lower limb two-dimensional sagittal kinematics from a monocular RGB camera during vertical jumps against a gold stand motion technology; they obtained a robust agreement with a RMSE below 3.2° across the hip, knee, and ankle trials (74). These studies deliver further evidence that markerless systems can be a practicable instrument for biomechanical research, extending beyond the laboratory, although they are yet to be proven for hands.

2.3 Dynamic hand gesture recognition

Automated action recognition is the method of identifying specific human movements by a machine (155). Specifically, hand gesture recognition has been used for different applications (156), (157). An often-used distinction embraces two main classes of hand gesture classification, including device-based and vision-based models (158). The first group attempts to solve the problem of recognition using sensor-based devices (159,160). However, these tools do not fulfil the human hand mobility naturalness, forcing users to carry and wear them while performing hand gestures and obstructing fingers physical appearance (161). Vision-based approaches (161) have the advantage of being unconstrained and non-obstructive, compared to instrumented device-based systems (158). Furthermore, the developments in the field of vision algorithms for hand gesture recognition have gained attention for their applicability to various different use cases, such as sign language interpretation and robotic applications (162).

Vision-based approaches for hand gesture recognition can be classified into models based on feature extraction and models based on observable features. Using feature extraction, algorithms can be loosely divided into appearance-based and model-based approaches (163). Appearance-based (or static) models are linked to a straight comparison of gestures against two-dimensional image features (163). These models use parameters derived from images. The most popular features include hand contours, colors, shape, optical flow, image edges, and other local hand features (161). Here, the hand pose does not vary during the gesturing time (164), as the algorithm aims to represent the observed features of the hand without any motion information (161). Some approaches have used filtering techniques based on morphological operations (165), applied to enhance the relevant details or remove holes and/or noise (166). Other studies have utilized Haar-like features (167) to detect edges or lines, with the drawback of tedious manual process and no motion information. Model-based approaches (16,152) employ a range of algorithms (e.g., OpenPose) to extract skeleton features as tracking tool to then assess the final gesture based on the keypoints extracted. The drawback of these approaches is the need for computationally intensive models that can limit the speed of the process and reduce its adoptability.

When hand gesture recognition is classified according to observational features, continuous video recordings are recognized based on temporal relationships, known as dynamic gesture recognition (Figure 13). Dynamic gesture recognition models observe and measure parameters such as orientation, trajectories, and speed (168), considering that spatial-temporal structure patterns characterise the motions. These motion patterns can be modelled along with multi-scale temporal trajectories, describing estimations of visible events. Since such large temporal trajectories can describe hand motion; dynamic gesture identification can be achieved by estimating the association or the length between such trajectories (169). The most challenging issue of dynamic gesture identification is the spatiotemporal mutability of different actions varying in length, appearance, and speed. These features make the task of dynamic gesture recognition more challenging compared to static gesture recognition.

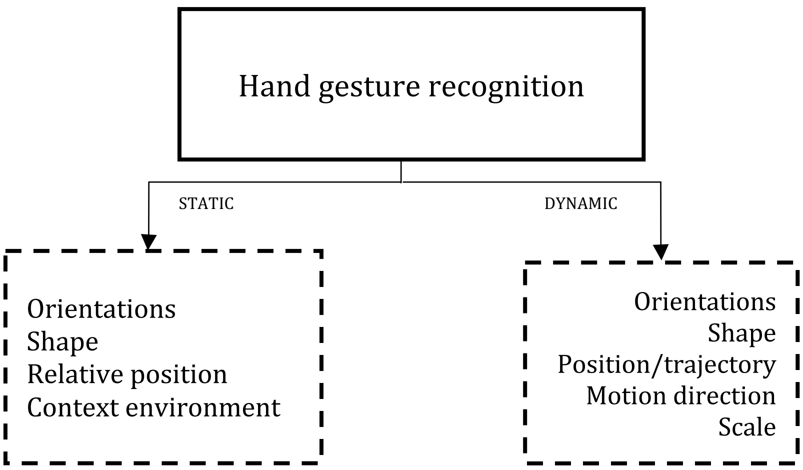


Figure 12: A flowchart to classify hand gestures into static and dynamic.

All of these investigations concentrate on action classification from video clips that are often previously trimmed to only consider a single action within a video sequence, even if video sequences are performed continuously during a recording (170). These continuous recordings have generated a volume of captured sequences that is constantly increasing. Starting from segmented sequences, existing studies (171) perform manual video segmentation to trim the videos into clips containing individual actions and expedite content extraction from each segmented clip. The problem with these methodologies is that they tend to disregard the long-term dependency of consecutive video frames within hand gesture actions. Furthermore, the process is labour intensive and not scalable. There is the need for a system that can efficiently recognise and segment actions within long video sequences.

2.3.1 Action classification

A number of traditional approaches have been explored for dynamic hand gesture classification, including dynamic time warping (172), the hidden Markov model (160), Gaussian process dynamical model (173). These techniques can observe dynamic hand gestures in terms of motion signature and trajectory. The drawbacks of these methods include the application to only short time frame tasks (172), or the adaptation of these probabilistic approaches to one state at the time (160), limiting their adoption with wide range of parameters (173). Furthermore, these traditional methodologies are difficult to generalize to different use cases.

Deep learning methods have aimed to reach better outcomes, considering the dynamic behaviour of hand motion, with the advantage of being more sensitive to learn rapid time-varying features. Taken individually, time, appearance, or space parameters, might not be sufficient to classify a gesture when considering long video sequences. To this end, conventional CNNs are often combined with other architectures to enable the model to learn the long dependencies that come from consecutive frames. While two-dimensional CNNs are generally adopted for appearance-based models, three-dimensional CNNs can be utilized for dynamic spatial and temporal features. However, 3D CNNs are inadequate to learn long-term temporal information (26). Molchanov et al. (174) offered a solution combining three-dimensional CNNs and RNNs. RNNs (142,143) extract parametric features by looking at current input and what the model has estimated from the inputs received previously. The aim of Molchanov's approach was to connect spatiotemporal features and then transfer them into an RNN. However, the spatial correlation knowledge was missing in the "pure" RNN phase. A few studies (25,26) have illustrated good performances by combining three-dimensional CNNs and LSTMs. However, these architectures have been adopted when both RGB and depth data are combined, and the question of whether RGB sequences alone can be fed into these architectures for action classification is left open.

2.3.2 Action segmentation

Numerous approaches have been presented to segment continuous clips from uncropped videos recordings. Kuehne et al. (175) proposed an end-to-end generative framework for video segmentation using the hidden Markov model for video segmentation and recognition of human activities, with the drawback of intensive manual labour. Ni et al. (176) presented an approach based on RNNs to perform sliding window detection to segment continuous actions. The issue with this methodology is linked to the identification of peripherical boundaries only, with no global overview of the temporal events.

To overcome these disadvantages, recent approaches have suggested making a distinction between gestural frames, when the action is taking place, and translation frames by merging both shape and spatiotemporal parameters. Such an approach has been presented by Wang (25) (Figure 13 and Figure 14).

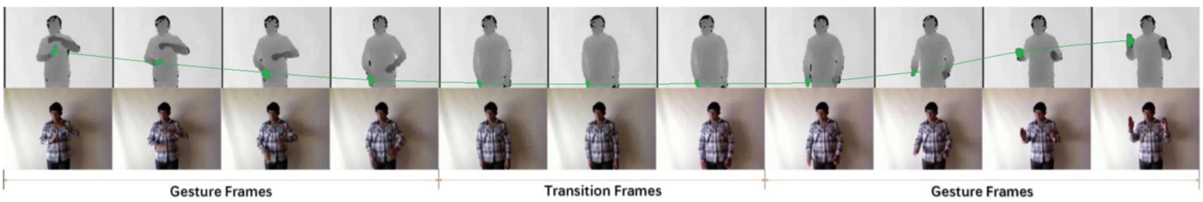


Figure 13: An example of continuous gesture sequence made of transitional and gestural frames. Reproduced with permission from (25).

This approach also takes in input RGB-depth data; an application for hand gesture classification and segmentation that utilizes RGB data is yet to be proven. Furthermore, this approach utilizes a binary classifier; a similar approach when multiple classes are involved is yet to be tested and evaluated.

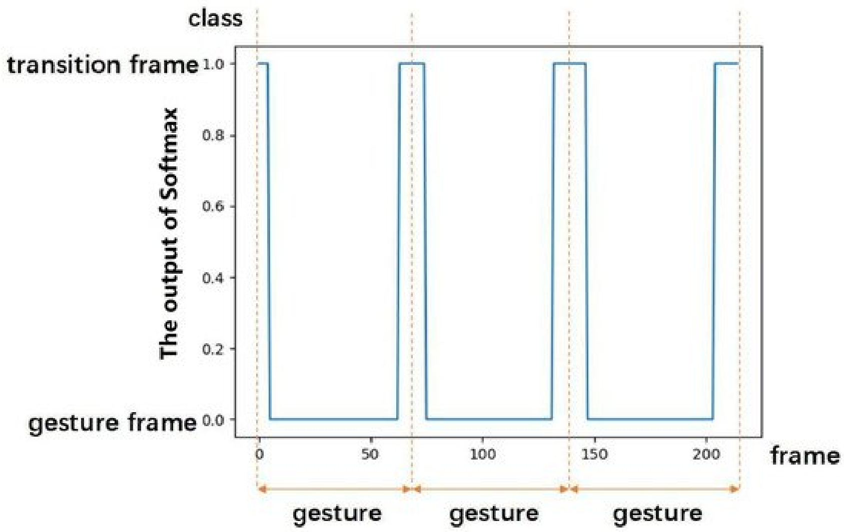


Figure 14: An example of the temporal segmentation results. Reproduced with permission from (25).

2.4 Discussion

Goniometric measures are widely adopted in clinic but have demonstrated high intra-reader variability. More accurate marker-based optoelectronic motion capture systems continue to be the gold standard metrics to capture hand kinematics but are difficult to adopt outside lab settings. The latest studies on deep learning architecture have enabled the capture of keypoints from monocular RGB cameras. These markerless approaches can support researchers to gain data from users in their natural environments. But while the computer-vision community keeps increasing the accuracy of markerless algorithms, the implementation of these models is generally outside the scope of classic hand biomechanics research.

In 2019, Seethapathi et al. (177) suggested that deep-learning-based human pose tracking algorithms did not prioritise the parameters that are important in the field of movement biomechanics. The application of markerless techniques in biomechanics settings may still be limited due to occlusion, and a lack of validated approaches. These challenges may be fixed in various modes. Several studies have suggested that a possible way to mitigate these limitations would be to ponder the temporal continuity in movement, introducing a gesture recognition component that would identify the joint dynamics and reconstruct the signal when tracking is lost. Recently, researchers have suggested that looking at the temporal continuity of gestures to reduce the impact of an error in keypoint detection models (160,168). This would enable automatic recognition of a motion path to enhance markerless tracking precision.

To address this action recognition task, initial research has focused on static hand gesture analysis (178,179). However, most recent works have offered a novel action recognition procedure that enlarges previous approaches by analysing the spatiotemporal continuity of the hand gesture (160,168). These techniques are known as dynamic gesture recognition and can deliver both classification and segmentation.

The usability of markerless tools to capture hand kinematics would also be a valuable tool in clinical research applications. For instance, enhanced comprehension of the mechanics that dictate the finger movement dynamics captured on impaired hands from their natural environment could provide support in current clinical practice.

Chapter 3 Rheumatoid arthritis

3.1 Background

Several studies have demonstrated how recent technology advancements in pose estimation applied to impaired populations can transform health care practice (20,28,180). The increased health care costs (181), the lack of specialized clinical staff (31), and drives in private health care systems to increase financial margins (182) are leading physicians to leave translational resource-intensive clinical practices and adopt alternative approaches to deliver care.

The rheumatology community, amongst other clinical specializations, has been adjusting to this developing landscape, welcoming possibilities that arise from data-driven digital clinical interpretation (183). As rheumatoid arthritis (RA) resides at the foundation of rheumatology practice, various improvements have been proposed to support RA management (47,184,185).

RA is a long-term, inflammatory symmetric polyarthritis in which the body's immune system mistakenly attacks the joints (186). It can cause persistent pain, damage, and long-term disability, especially in the hands (187). The condition usually starts in the small joints of the hands and later spreads to involve the larger joints (188). Therefore, for the majority of people living with this condition, hands are often affected by swelling, weakness, and restricted mobility, resulting in loss of function that can bring significant challenges, such as physical suffering and diminished quality of life (186).

Given the degenerative nature of RA, early diagnosis is important for preventing the progression of inflammation and deformities, with gradual joint damage that can steer to mobility decline (187). There are several validated assessment methods used by clinicians to investigate symptoms of RA (39). Van der Heijde et al. introduced what is nowadays considered the leading metric to assess the disease activity in RA, known as the Disease Activity Score using 28 joints (DAS-28) (189).

Using DAS-28, RA is monitored through routine visits at settled periods (190), usually occurring every 3 to 6 months. However, disease severity and signs of illness progression vary hugely between appointments, and it has been evidenced that fixed intervals may not detect the crucial time points of such symptom aggravation (181). Frequent monitoring can open up to better disease management, decreasing health care costs, while maximizing the effects of therapies and reducing disability (48). However, for RA frequently in-person consultations attendance is low (183), which generate worsening symptoms (34), and increases health care costs.

To overcome these limitations, remote monitoring tools have recently emerged as valuable instruments for supporting RA management. These approaches (191) have been using patient-initiated endpoints to offer the opportunity to trace symptom severity in everyday settings. Despite their ambition, evidence suggests huge variability in engagement with such remote monitoring solutions due to the lack of objectivity when the clinical endpoints are traced (183).

In parallel with these solutions, to contain disease activity and decrease hospitalizations, the National Institute for Health and Care Excellence (NICE) RA pathway (192) recommends a 3-pronged management plan; i) non-pharmacological management (hand exercise programmes), ii) drug treatment (disease-modifying antirheumatic drugs (DMARDs) and rituximab) and iii) surgical treatment. Although medications improve hand symptoms, the range of motion (ROM) and muscle function are not regained (193). This is because DMARDs and biologics (a form of pharmaceutical treatment) do not reverse muscle wasting. To address this issue, hand exercise programmes are a standard component of RA management practice.

Specific hand activities have been developed, demonstrating an enhancement in joint ROM for individuals with RA (194). These hand exercise programmes have been indicated to be effective in the management of RA; however, a systematic review (183) suggested that patients do not adhere to these programmes. The main reason is linked to those systems not being connected with patients' therapy regime and disease activity, and a lack of objective outcome tracking.

As hand ROM is a strong indicator of disease activity (195,196), and a crucial component in the management of RA, a few attempts to measure hand ROM for this population have emerged. This chapter describes objective hand tracking methodologies used in RA and their limitations in adopting such solutions in home-based settings. Notwithstanding their end goals, these objective tracking methods have restrictions. A great deal of effort is required for both physicians in evaluating the data, and RA patients in discovering how to interact with the system (197). Simblett et al. suggested potential interest of RA patients in a monitoring platform that can easily capture disease activity and objectively track joint ROM during at home-based exercise programmes, limiting the burden on both clinicians and patients (198).

3.1.1Epidemiology and costs

The National Audit Office estimates that approximately 580,000 adults in England currently have RA, with a further 26,000 new cases diagnosed each year and approximately 1% of the UK adult population affected (188). The condition affects 0.6% of the US population (199), with 1.3 million American adults estimated to be affected in 2005. By 2007, the number increased to 1.5 million adults with RA in the United States (200). In Europe, RA has a prevalence of up to 1.1%, while the annual incidence varies between about 20 and 50 cases per 100,000 inhabitants (32). In general, the prevalence of RA in the population of developed countries ranges approximately from 0.5% to 1.1% (200). Mortality hazards are usually 60% to 70% higher in patients with RA compared to the general population (201), and the disease is three times more frequent in women than men (40). In the UK, the prevalence of RA is 1.16% in women and 0.44% in men, increasing with age to 5% in those aged over 55 years (202).

The lost productivity associated with RA is substantial because of the progressive nature of the condition (203). Many individuals report missing work or choosing not to work because of disease-related disabilities (203). Approximately 20% to 70% of individuals with RA who were working at the inception of their condition had to leave work after seven to ten years because RA resulted in disability (203). As a consequence, in developed countries the indirect cost of RA has been estimated to be nearly three times higher than the cost of treating the disease (204).

Based on a report by the National Rheumatoid Arthritis Society in 2010, the overall cost of RA to the UK economy was almost £8 billion per annum, with NHS expenditure totalling approximately £700 million (205). In parallel, based on 2005 U.S. Medicare/Medicaid data, the total annual societal costs of RA (direct, indirect, and intangible) reached $39.2 billion (206). These costs included direct ($8.4 billion) and indirect ($10.9 billion) costs, intangible costs due to quality-of-life deterioration ($10.3 billion), and increased costs due to premature mortality ($9.6 billion) (206).

3.1.2 Aetiology

The detailed aetiology of RA is unknown, but advances in molecular research have attributed 50% of the risk of developing RA to genetic factors (207). The onset of the condition starts when the T-cells of the immune system infiltrate the connective tissue that lines the inside of the joint capsule, called the synovium, leading to synovial joint damage that results in hypertrophy (enlargement of the tissue) and inflammation of the local area (208). The joint damage occurs because of the inflamed synovial membrane, also known as erosive synovitis. In particular, if the inflammation goes unchecked, it can damage the cartilage and, in some cases, can have extraarticular involvement (209).

Over time, this loss of cartilage makes the joint spacing between bones become smaller. As a consequence, the joints affected can become unstable, painful, and lose their mobility (210). The resulting inflammation (Figure 15), causes the inside of the joints to thicken, leading to eventual degenerative joint destruction (39).

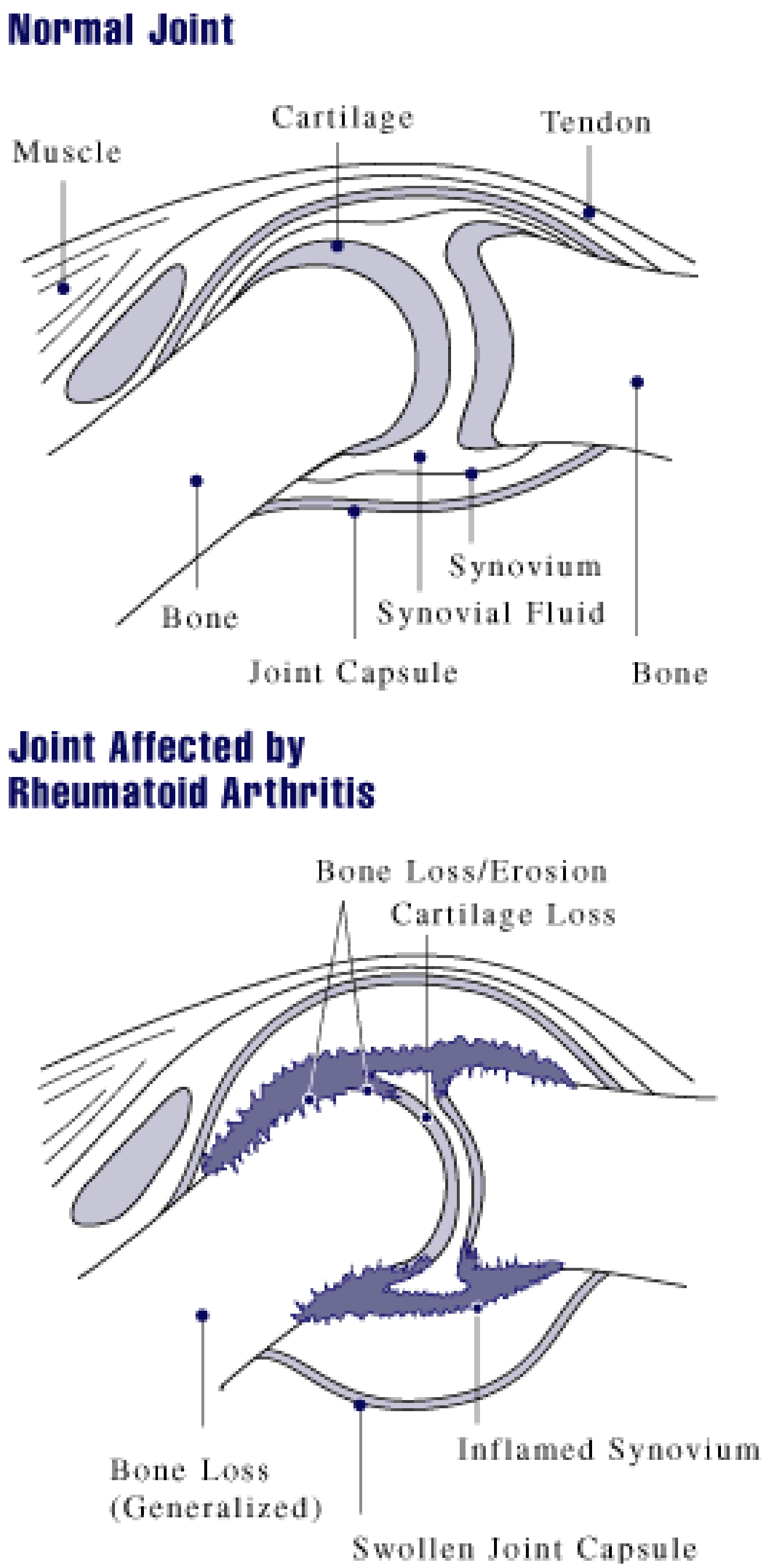


Figure 15: Rheumatoid arthritis joint breakdown. Reproduced with permission from (61).

Inflammation, therefore, causes the joints to degenerate, which leads to progressive joint damage that cannot be reversed (39). Deformities arise because of joint cartilage being eroded, which can then extend into the bone cortex. Using body compensatory mechanisms may also result in high joint forces (211) generating deformities. In the hand, deformities can reach all articulations, causing partial dislocation of a joint, also known as subluxation (212), and deformities in the metacarpophalangeal (MCP), interphalangeal (IP) joints, and the wrist (40).

The common forms of deformities in RA occur due to volar subluxation of the proximal IP joints and ulnar deviation at the MCP joints (213) or swan-neck and boutonnière deformities for instability at the MCP and IP joints of the fingers and thumb, respectively (214) (Figure 16). These deformities can affect some or all the joints of the hand (Figure 17).

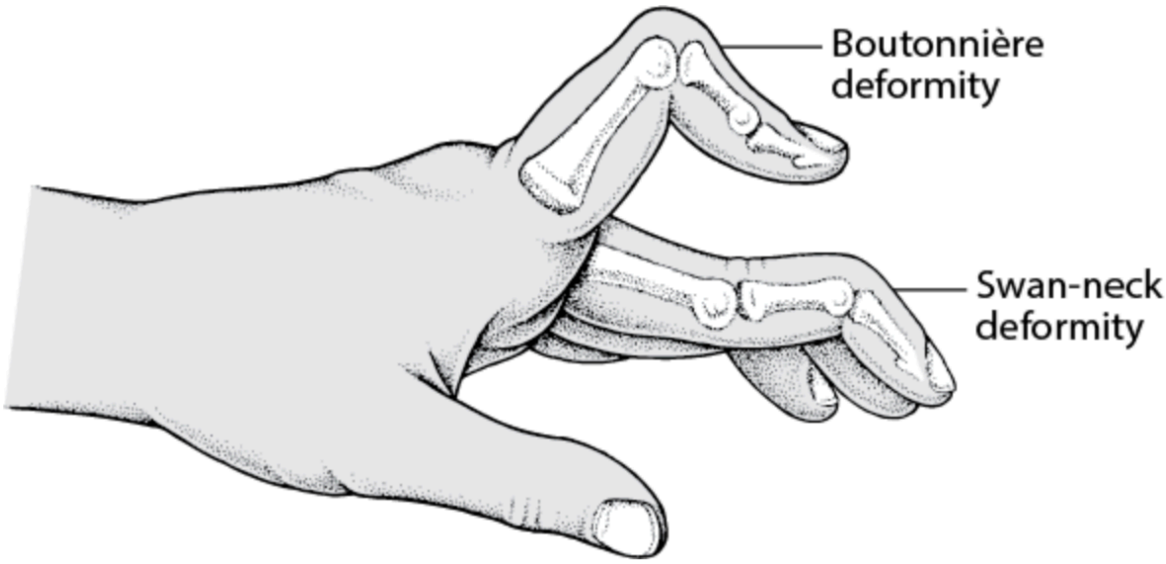


Figure 16: Illustration of hand deformities due to rheumatoid arthritis showing swan-neck deformity and boutonnière fingers deformity. Reproduced with permission from (215).

(A) (B)

Figure 17: (A) Two hands showing bilateral metacarpophalangeal (MCP) synovitis evolving to boutonnière deformity. (B) Two hands illustrating bilateral ulnar deviation, swan-neck deformity, fully developed boutonnière deformity. Reproduced with permission from (216).

3.1.3 Clinical assessment

In 1969 Ansell presented a study, proposing how a standard hand examination should be conducted on individuals with RA (217). In this paper it was suggested that the examination of the hand should start at the wrist, to investigate whether it is fixed in palmar flexion. Interrogation should include the site of pain, the presence of stiffness and its duration, symptoms of altered sensation, and which is the dominant hand. Then, each of the MCP, PIP, and DIP joints should be examined for swelling, fluid accumulation, and bony enlargement. According to Ansell, the range of movement at each joint should then be recorded, together with any deformity. The limitation of this assessment is that it relies solely on visual examination, which is linked to the expertise of the assessor (218). In 1987, to remove the hierarchy of certainty around diagnosis, the American Rheumatism Association developed a table that outlined the seven major criteria for classifying RA around diagnosis (219) (Table 1).

Table 1: 1987 American College of Rheumatology Criteria for rheumatoid arthritis classification. Table reproduced with permission from (219).

Criterion	Definition
Morning Stiffness	Morning stiffness in and around the joints, lasting at least 1 hour before maximal improvement
Arthritis of 3 or more joint areas	At least 3 joint areas simultaneously have had soft tissue swelling or fluid (not bony overgrowth alone) observed by a physician. The 14 possible areas are right or left PIP, MCP, wrist, elbow, knee, ankle, and MTP joints
Arthritis of hand joints	At least 1 area swollen (as defined above) in a wrist, MCP, or PIP joint
Symmetric arthritis	Simultaneous involvement of the same joint areas (as defined in 2) on both sides of the body (bilateral involvement of PIPs, MCPs, or MTPs is acceptable without absolute symmetry)
Rheumatoid nodules	Subcutaneous nodules, over bony prominences, or extensor surfaces, or in juxta-articular regions (near a joint), observed by a physician
Serum rheumatoid	Demonstration of abnormal amounts of serum rheumatoid factor by any method for which the result has been positive
Radiographic changes	Radiographic changes typical of rheumatoid arthritis on posteroanterior hand and wrist radiographs, which must include erosions or unequivocal bony decalcification localized in or most marked adjacent to the involved joints (osteoarthritis changes alone do not qualify)

A range of studies over the last three decades has looked at conventional radiographs of the hands and feet to assess structural joint damage due to RA (220). These radiographic measures have been adopted due to their ability to provide an objective marker of disease activity and assess any improvements or failures of treatments. However, it has been demonstrated in several studies that these techniques are not sensitive enough to detect changes early in the disease process (221). This is because erosions may only become visible up to two years after the onset of disease, and soft tissue involvement may not be detected at all (222). Considering these limitations, alternative methodologies that can provide a detailed and early quantification and detection of disease activity have been investigated.

Currently, RA patients are assessed and examined in outpatient clinics following the treat-to-target guidelines provided by the European League Against Rheumatism (EULAR) (190). The guidelines state that the DAS-28 should be captured and stored frequently, every month for patients with high/moderate disease activity, and every 3-6 months for patients with low disease activity scores or who are in remission. The DAS-28 involves four domains (223) including a clinician-reported swollen joint count (SJC), a clinician-reported tender joint count (TJC), a global measure of symptoms, and a biomarker of inflammation.

Joint tenderness is the presence of pain in a joint at rest with pressure or on the movement of the joint (224). To obtain the TJC, the examiner documents which joints the patient indicates are painful on palpation with enough pressure to blanch the nail bed of the examiners thumb and index fingers (225). Joint swelling is recorded as soft tissue swelling that is detectable along the joint margins, when a synovial effusion is present (224). Fluctuation, defined as significant disease variation exceeding the standard deviation of the previous scores (226), is a characteristic feature of swollen joints. Large fluctuation influences the range of joint movement, particularly in the small joints of the hand (224). To assess the SJC, the examiner documents which joints have palpable soft tissue swelling, excluding joints affected only by deformity or bony hypertrophy (225).

Establishing joint swelling is considered in literature (227) to be the only objective and the most crucial component of the DAS-28 for two reasons: (1) joint swelling is a key predictor of future damage within a joint and (2) joint swelling is linked to joint range of movement. It is therefore essential that the swollen joint component of the DAS is accurately assessed. Excessive swelling can cause a reduced joint ROM. To address this issue several techniques have been proposed to objectively measure joint ROMs. The introduction of objective measures of disease activity presents a landmark change in the management of RA, and it is now routinely measured in clinic visits (228) using goniometric assessment, to assess ROM in heathy and swollen joints. The clinician reported SJC and TJC considers 28 joints, including the MCPs and PIPs of the fingers and the thumbs, IPs of the thumbs, wrists, elbows, shoulders, and knees (Figure 18).

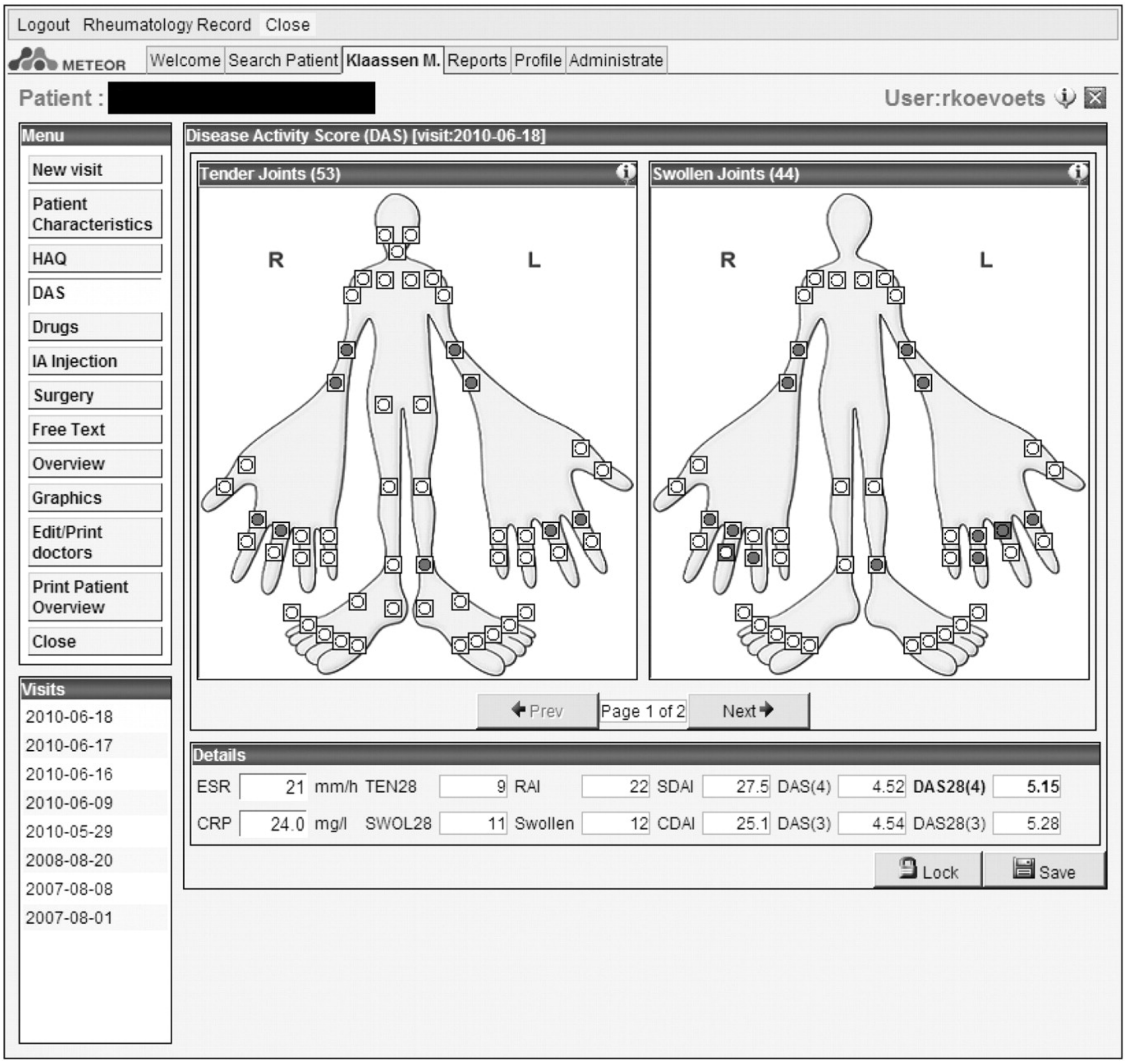


Figure 18: Illustration of tender and swollen joints counts of the Disease Activity Score 28 (DAS-28) representing a fictitious case. Reproduced with permission from (229).

The other two components of the DAS-28 are a global measure of symptoms and a biomarker of inflammation. The global measure of symptoms is tracked using a patient's global health assessment (GH), which involves pain evaluation using a 100 mm visual analogue scale with 0=best, 100=worst (228). A biomarker of inflammation estimates the blood laboratory parameters of an acute phase reactant, such as erythrocyte sedimentation rate.

Once all the parameters are obtained, the DAS-28 calculation reports a score between zero and nine, with scores below 2.6 reflecting excellent disease control, while scores >5.1 indicate severe disease activity. The four items, including the TJC measure ($\text{TJC} \in [0, \dots, 28]$), the SJC measure ($\text{SJC} \in [0, \dots, 28]$) and the global health (GH) pain measure (as a visual analogue score in mm $\text{GH} \in [0, \dots, 100]$), combined with the c-reactive protein (CRP) measure (CRP in $\text{mg/L} \in [0, \dots, 300]$) in equation 3.1 or with the Erythrocyte Sedimentation Rate (ESR) measure (ESR in $\text{mm/hr} \in [1, \dots, 300]$) in equation 3.2 obtained from the blood tests, are entered into a formula and are calculated as follows to obtain a dimensionless DAS score:

$$DAS28_{CRP} = 0.56\sqrt{TJC28} + 0.28\sqrt{SJC28} + 0.014GH + 0.36\ \ln(CRP + 1) + 0.96 \tag{3.1}$$

$$DAS28_{ESR} = 0.56\sqrt{TJC28} + 0.28\sqrt{SJC28} + 0.014GH + 0.70\ln(ESR) \tag{3.2}$$

Equations 3.1 and 3.2 also include the Ritchie articular index (230), and cut-off values of low and high levels of disease activity derived and published (231). Finally, the appropriate cut off point for remission, low, high and very high disease activity has been measured (231), and was most recently validated statistically (232); for instance, the remission cut-off value over a longitudinal study was reported to be 2.6.

3.1.4 Limitations

Multiple clinical trials (233) over three decades have shown that the best outcomes are achieved when patients are treated to obtain the target of extremely low disease activity. To succeed in this, RA patients need frequent monitoring. The best outcomes have been observed in studies that reviewed disease activity monthly (234,235), titrating medication accordingly. Review appointments should take place a month after treatment to enable patients to be transferred onto alternative treatment pathways. However, the 2009 National Audit Report revealed that only 10% of RA patients were receiving adequate follow-ups (236).
Another problem is the adoption of the DAS-28 in rheumatology practice. The DAS-28 score has demonstrated great clinical value in evaluating, monitoring and treating individuals with RA (218). However, only a small percentage of rheumatologists have incorporated these tools into their standard, everyday clinical practice (237). This is due to the time required to administer a questionnaire, assess the patient's joint pain and swelling, score the results and record the information in a readily retrievable format (238).

Individuals living with RA undergo frequent symptom fluctuations, with an increased level of joint inflammation in between clinical appointments and exacerbated illness signs, known as flares (239). These flares are intense episodes of the disease manifestation at an unpredicted point in time. However, clinical consultations happen at fixed points in time. Furthermore, these consultations rely upon the history of flare episodes, which is subject to recall bias and difficulty in summarizing symptom intensity objectively.

RA treatment using medications that suppress the immune system are titrated according to disease severity. Treatments include painkillers, non-steroidal anti-inflammatory drugs, steroids, and disease-modifying drugs (192). Each patient typically undergoes an evaluation of each medication class or combination of medications starting with the least aggressive and then escalated upwards until the disease becomes manageable (low disease activity/symptoms) or ideally in remission, i.e., treat-to-target (233). However, even then, disease flare-ups often occur; these are poorly documented and harm the therapy regime, which can cause further pain and distress.

3.2 Alternative tools to capture clinical endpoints

3.2.1 Smartphone-based technologies

Remote measurement tools have appeared as valuable instruments to promote health management, and there are increasing indications to confirm the cost-effectiveness of such interventions. They represent a fast-growing field for the provision of care.

There has been a growing appetite to adopt such remote solutions into the clinical care of RA to provide a device for patients to be more actively connected in their disease. Many techniques have emerged to increase the usability of the DAS-28, using remote monitoring. Amongst the different components of the DAS-28 score, biomarker levels can easily be obtained using commercially available home CRP testing kits (43). In 2007, Figueroa et al. demonstrated that, while the SJC of patients and physicians correlated poorly, a patient's self-assessment of joint tenderness is reliable with physician recognition (238). Several studies have proposed computerized questionnaires to capture traditional patient symptom severity scores, allowing the data from all previous visits to be readily available using mobile applications.

A review on the mobile applications for the management of RA has identified a total of 19 tools able to run on Android or iPhone Operating Systems (iOS) for symptom assessments in RA (191). In this systematic review, Grainger et al. illustrated that several applications did not adhere to standard guidelines for how to capture disease activity and did not involve clinical experts during the development of these tools, leading to uncertain outcomes. However, it has also been shown that when clinical teams are involved in the co-design of mobile applications, these solutions can have a positive impact on the health outcomes of patients with RA (183).

An example of a mobile technology co-designed with clinical team is the approach of Austin et al. (42). They conducted a study named Remote Monitoring of Rheumatoid Arthritis (REMORA). During their research, they integrated electronic health records to detect daily symptoms, including flares. The objective of this investigation was to demonstrate the acceptability of reporting symptoms using a smartphone app and found that this approach was adopted confidently by people with RA and their clinical team, with a high degree of engagement. Nevertheless, visible signs of disease progression, which are considered fundamental to assess the disease activity (240), were still lacking. In another investigation, the disease activity was captured collecting both objective kinematic data, using an accelerometer, and symptom reporting, such as patient-reported questionnaires and digitally recorded joint counts (43). As this solution also contained data on the range joint ROM, the aggregate of these two types of data could predict in-clinic RA activity. However, this level of input for remote monitoring solutions is unusual.

Figueroa et al. established the importance of a mechanism to remotely assess the ROM, as it is linked to how many joints are swollen (241). The assessment of mobility, particularly hand mobility, can also potentially serve as an early indicator of change in disease activity and thus allow timely adaptation of patient management procedures (41). This early quantification and detection of disease activity would be particularly important to allow personalized treatment regimens in remote monitoring applications and amend subjective patient-reported outcome measures (234). Objective quantification can provide an important step towards understanding hand movement disorders and evaluating the effect of possible interventions (242). However, multiple reviews have suggested a lack of objective data in the apps available (31,191,243,244).

3.2.2 Physical-therapy management

It has been estimated that the application of pharmacological treatment in RA raises direct healthcare costs by 300% (245), while cost-efficiency remains questionable. Moreover, muscle strength and joint ROM are not immediately increased when using drug treatment (193). Amongst the non-pharmacological management approaches, hand exercise programmes are low-cost approaches that have proved to enhance joint mobility (195), and may even enhance the effects of medication (246).

Several approaches have been proposed for exercises to improve hand mobility for RA (184,195,247). Various types of exercise are employed to address different aspects of the

RA patient experience, with the majority reporting beneficial responses (194). Hammond et al. (193) evaluated nineteen trials to assess the effectiveness of functional hand exercise programmes in RA. They used the PEDro scoring system (248), consisting of 11 criteria to determine the validity of a study. The top-scoring study, introduced a generic upper limb programme, named the Education, Self-Management, and Upper Extremity Exercise (EXTRA), for people with RA (249). The study with the next greatest score developed a hand and wrist specific exercise programme for individuals with RA, named the iSARAH programme (246). The iSARAH hand exercise intervention consisted of seven mobility exercises and four strength exercises against resistance (Figure 19). The mobility exercises included MCP flexion/extension, tendon gliding, finger radial walking, wrist circumduction, finger abduction/addition, hand behind the head, hand behind the back. The strength exercises included eccentric wrist extension, gross grip, finger adduction, and pinch grip. The incorporation of these hand exercise programmes in mobile applications has been explored in different studies (250), merging hand exercises and qualitatively questionnaire to support self-management interventions, while enhancing both adherence and long-term effectiveness.

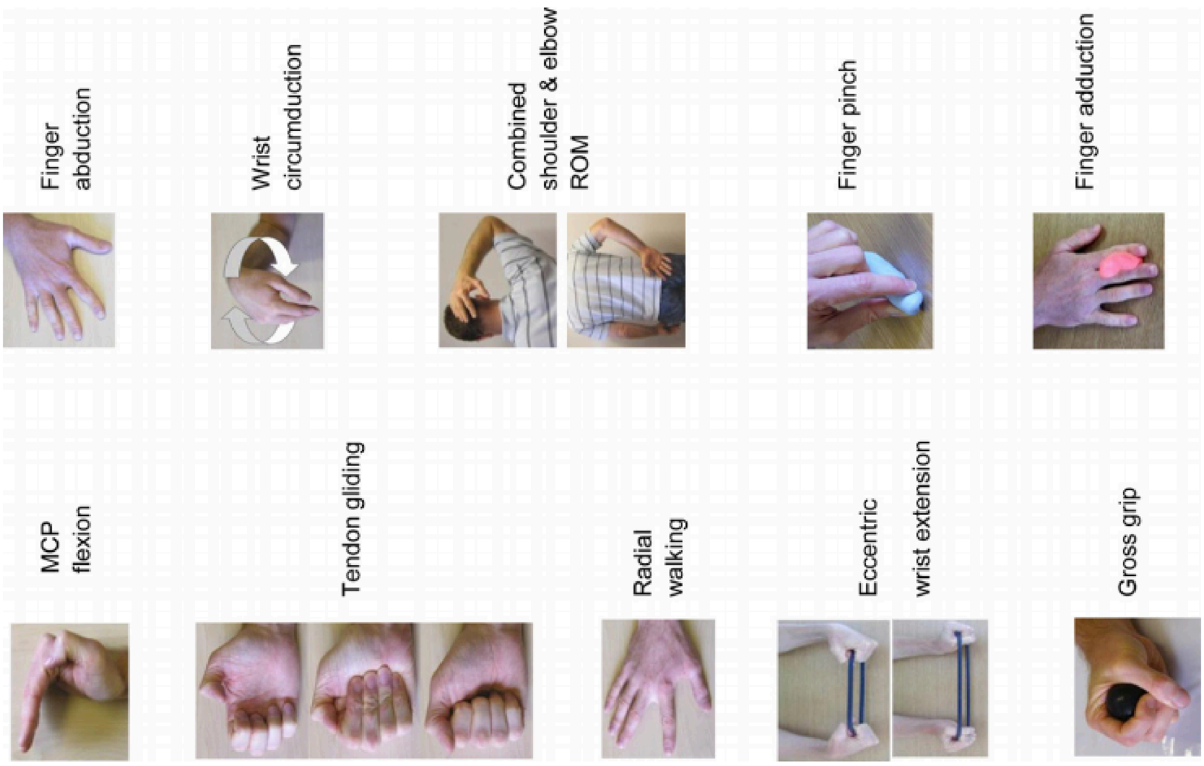


Figure 19: The Strengthening and Stretching for Rheumatoid Arthritis of the Hand exercise programme. Reproduced with permission from (251).

A systematic review of eight studies on hand exercises for RA published between 2000 and 2014 concluded that remote hand exercises improve hand ROM (243). As the range of movement is directly linked to the disease activity score, it is important to assess objectively the mobility range of RA individuals when these techniques are embraced to provide a more targeted intervention. As result, disease activity scores when hand exercises are adopted, have demonstrated significant improvements in the short term (at 2.5 (249), 3 (252), and 4 (246) months).

Despite their purpose, research suggests large variability when RA patients engaged with such remote solutions with low (11%) levels of adherence (183). In several investigations this was linked to the difficulty in fulfilling a qualitative survey (183,253,254), but mostly it was reported that individuals perceived low clinical value as the data were not objective and not connected with their illness intensity and disease activity (198). Moreover, studies have demonstrated that these programmes demand a high effort for the sufferers and the health care team involved (195,255). However, when patients are given feedback (197), or as soon as the data are linked with the clinical team to also consider their therapy regime (256), adherence to these interventions can be improved (256).

3.2.3 Marker-based objective technology-based assessments

Clinicians and researchers have acknowledged the importance of objective measures of hand mobility to benefit RA assessment (39). Smolen et al. (257) also recognized that physicians can reliably evaluate the condition by monitoring and recording hand motion daily. Similarly, Majithia and Geraci (258) discussed how it is vital to diagnose RA at early stages to prevent the development of small joint erosion. To address the need for an objective component that could track remotely the range of movement from RA patients a few techniques are presented.

Goniometry has shown short-term efficacy of the treatment of synovitis in RA (259), and improved treatment interventions and enhanced quality of life scoring (260) when used in combination with closed monitoring to deliver targeted treatment interventions. However, the process is manual and demands a qualified therapist, which makes it repetitive and cumbersome. Furthermore, as discussed in Chapter 2, the goniometric assessment presents reliability concerns due to inter and intra-reader inconsistencies (120).

Marker-based optoelectronic motion capture technology has also been used to obtain accurate measurements of joint mobility for this population (261,262). These methods can correctly assess the joint ROM against values of a reference (healthy) population to determine the extent of disability (263). One study reported mean reductions of up to 17% in hand ROM in the early stages of the disease (i.e. within 7–12 months of diagnosis) compared with age- and gender-matched healthy individuals (264). A study reported that RA patients had finger ROM deficits of 17-28% after two to four years, compared with healthy volunteers, rising to 35-49% after eight years (265). The most widely accepted standard values that have been used as an indication of disease activity describe a loss of mobility of 20-30% on average for RA patients. Epps et al. (266) suggested the implementation of a global joint range of movement score to catch the significance of hand movements in RA. The results of the ROM for a healthy population compared to an RA population for finger joint movement are reported in Table 2.

Table 2: The metacarpophalangeal (MCP) joint, proximal (PIP) joint, and interphalangeal (IP) joint ranges of motion (ROM) for healthy and rheumatoid arthritis (RA) participants.

Joint	Finger	Healthy ROM	RA ROM (flexion/abduction)
MCP	Thumb	0° to 55° (267)	below 25°(268)
	Index	0° to 90° (267)	below 55° (268)
	Medium	0° to 90° (267)	below 60° (268)
	Ring	0° to 90° (267)	below 40° (268)
	Little	0° to 90° (267)	below 50° (268)
PIP	Thumb	0° to 80° (267)	below 60°(268)
	Index	0° to 100° (267)	below 40°(268)
	Middle	0° to 100° (267)	below 40°(268)
	Ring	0° to 100° (267)	below 40° (268)
	Little	0° to 100° (267)	below 40° (268)
Finger Intersects (web-space)		25° to 50° (267)	below 20° (268)

The problem with optoelectronic systems, as discussed in Chapter 2, is the difficulty in adopting them outside laboratory settings. Therefore, over the past few years several markerless techniques have been tested to capture clinical endpoints from uncontrolled environments.

3.2.4 Markerless assessments

As discussed in Chapter 2, joint ROM may also be assessed using digital photogrammetry. A preliminary approach to quantitatively measure rheumatic hand ROM from photographs was proposed by Highton et al. (269) in 1996. In their work, they used images of hands illustrating which anatomical features should be used (Figure 20).

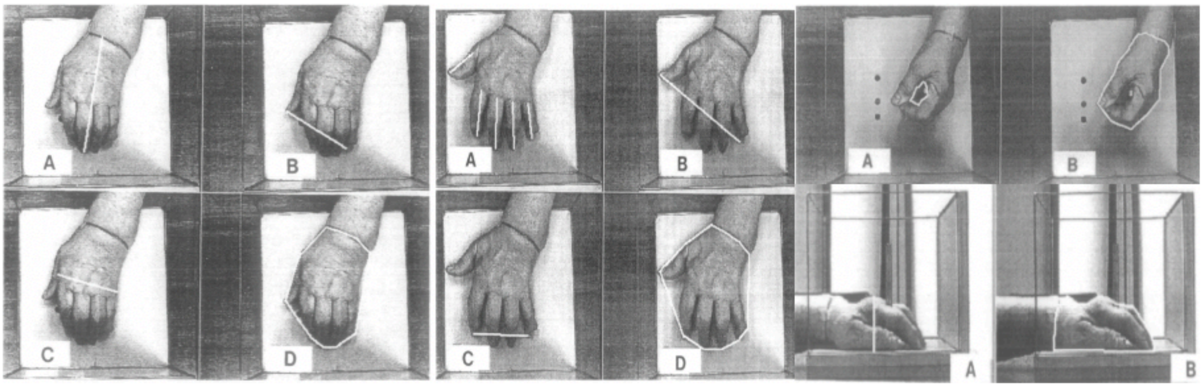


Figure 20: Images of dorsal and lateral view of closed, spread, and open hand of the twelve measures used by Highton et al. to assess the rheumatic hand. Reproduced with permission from (269).

Although this method illustrates clear parameters to look at when assessing a rheumatoid hand, Meals et al. (270) demonstrated that digital photogrammetry has limited effectiveness compared to goniometric assessment, which in itself has shown vast inter-rater variability.

To address the need for more accurate approaches to measurement, sensors and motion sensing devices have been explored. Hamy et al. (43) presented an objective way to capture motion parameters using gyroscopes and accelerometers in iPhones. While their study could only be used to assess wrist mobility, these investigations also suggested the need for an objective, more patient-centric digital endpoints data collection to track the progress of RA.

Similarly, Lima et al. (271) used a Leap Motion Controller™ sensor to estimate hand angles; however, they got high errors when the two methods were compared. This is emphasised by the recent work from Ganguly et al. (272). They presented a comparison between the gold-standard optoelectronic motion capture system and the Leap Motion Controller™, suggesting that the Leap Motion is not suitable for hand motion capture in clinical settings.

Most recently, Phutane et al. (41) presented a study to monitor hand motions in RA patients using markerless radar technology. The study illustrated the potential for the development of a markerless recording technique that can support the characterization of hand movement, showing that the system could be used to acquire measurements from a variety of movements observed in activities of daily living. However, the system is not scalable in a home-based setting and requires a high level of human feature manipulation.

In recent years, the development of deep learning algorithms has led to significant advances in video-based markerless hand tracking, as discussed in Chapter 2. Some of these approaches have been extended to objectively track RA. Cejnog et al. presented a framework for automatic hand ROM evaluation of RA patients using an Intel RealSense® SR300 depth sensor (40). They used a Convolutional Neural Network (CNN) named Pose-REN (273) to estimate the 21 points of reference. The flexion/extension angles were obtained by extracting the vectors between the adjacent joints for the MCP, PIP and DIP, while for abduction/adduction the opening angle was computed as the angle between the midpoint of the MCP joints of both fingers and each PIP joint. This study demonstrated that hands with RA could be automatically discriminated from those of healthy participants. The angles at the joints could be used as an indication of current movement capabilities from a simple movement cycle and this was enough to distinguish RA patients from the control group.

Even if the solution suggested by Cejnog et al. is novel and more scalable than a laboratory-based solution, depth cameras still have several limitations when adopted in home-based settings (e.g., ambient lightening saturating the sensor), as discussed in Chapter 2. All the described technologies, represent an effective yet limited way of capturing clinical endpoints objectively. However, they still require highly engineered and customised features to be adopted in a home-based setting.

3.3 Discussion

Movement dysfunctions have been often reported in RA patients (195). The impaired status of RA patients can be explained by the degradation of the joints, which creates swelling that results in a decreased joint range of movement, as reported in the EULAR recommendations for the management of RA (190), particularly in the small joints of the hands (188). There is a need for embracing novel methodologies to support the diagnosis and management of RA. By doing so, remote monitoring apps, such as those discussed in the systematic review by Grainger et al. (191), play an important role, but lack objectivity.

Since the hand is a good indicator of disease activity and is the part of the body predominantly affected by this autoimmune disease, manifold interventions have been designed to maintain and improve mobility with hand exercise programmes. To reach the integration of a system that can remotely and objectively support the management of RA, some techniques have been presented (40,41,43), but demand delicately adjusted parameters.

The field of remote management of RA continues to strive to provide a better and optimized treatment decision. The new era of markerless hand pose estimation from monocular RGB cameras has demonstrated great improvements in other clinical domains and can be an opportunity to reconstruct remote management strategies for RA.

Chapter 4 Comparison of contact and non-contact measures to track hand kinematics in healthy participants

4.1 Introduction

Optical marker-based motion tracking systems have been used to measure motion parameters and are categorised based upon their working principle, dividing them into marker-based and markerless (65). Marker-based motion capture systems work through active and passive tracking, for instance, by attaching reflective markers while using infrared (IR) cameras to record motions (55). Passive optical marker-based settings are considered the *gold standard* to measure human movement in the field of biomechanics (274). However, conventional marker-based motion capture systems are expensive, localised to the laboratory, not easily accessible to the broad population, and are difficult to adopt in medical environments (275,276).

In 2020 a survey (277) conducted on affiliates of the Academy of Orthopaedic Physical Therapy discovered that 57.9% of patients and clinicians were using video content captured from their personal devices to qualitatively assess human motion during orthopaedic physical therapy. In parallel, Owoeye et al. (278) expressed the need for portable, scalable, and quantifiable solutions that could be added efficiency to current resource-intensive clinical procedures.

The work presented in this chapter was partially funded by Innovate UK Project No: 75908.

To extend the validity of OpenPose to the clinical population, Sato et al. (20) assessed the walking cadence of the two-dimensional sagittal keypoints assessed with OpenPose against a traditional motion capture for people with Parkinson's disease. More specifically, their approach aimed to calculate the gait distribution and assess the periodicity of each sequence. The study reported good agreement between the two methodologies. However, their investigation incorporated only two people with Parkinson's and needed hand-crafted parameters. Therefore, Kidziński et al. (28) extended the pipeline presented by Sato et al. (20) for 1026 people with cerebral palsy.

The objective assessment of finger kinematics is fundamental to enhance the knowledge of hand mobility in both healthy and impaired populations. However, from the previously described investigations, it is not clear if these studies can be adapted to address the specific needs of finger kinematics, where relative segment motion is on a much smaller scale than the relative segment motion during human gait; although, the angular ROM of the joints in the hand are often higher that those seen in gait. Therefore, this chapter aims to compare the three-dimensional kinematics obtained with a gold standard marker-based optoelectronic motion capture system (Qualisys AB, Gothenburg, Sweden) against two-dimensional hand kinematics obtained executing OpenPose on frames acquired from a monocular RGB camera.

4.2 Materials and methods

4.2.1 Experimental setup

The protocol was approved by the Imperial College Research Ethics Committee (ICREC). Twelve healthy volunteers (eight female, four male) participated in this experiment (Figure 21). Participants were asked to attend a single session of recording of hand kinematics at the Upper Limb Motion Analysis Laboratory (White City Campus, W12 0BZ). Upon arrival, participants were briefed on the project, guided through a review of the participant information sheet, asked to sign the consent form, and informed of the set of sequences to perform. Written informed consent was obtained from each participant. Participants were visually supported by a PowerPoint presentation that guided them through the hand exercises to be performed with both the right and left hands.

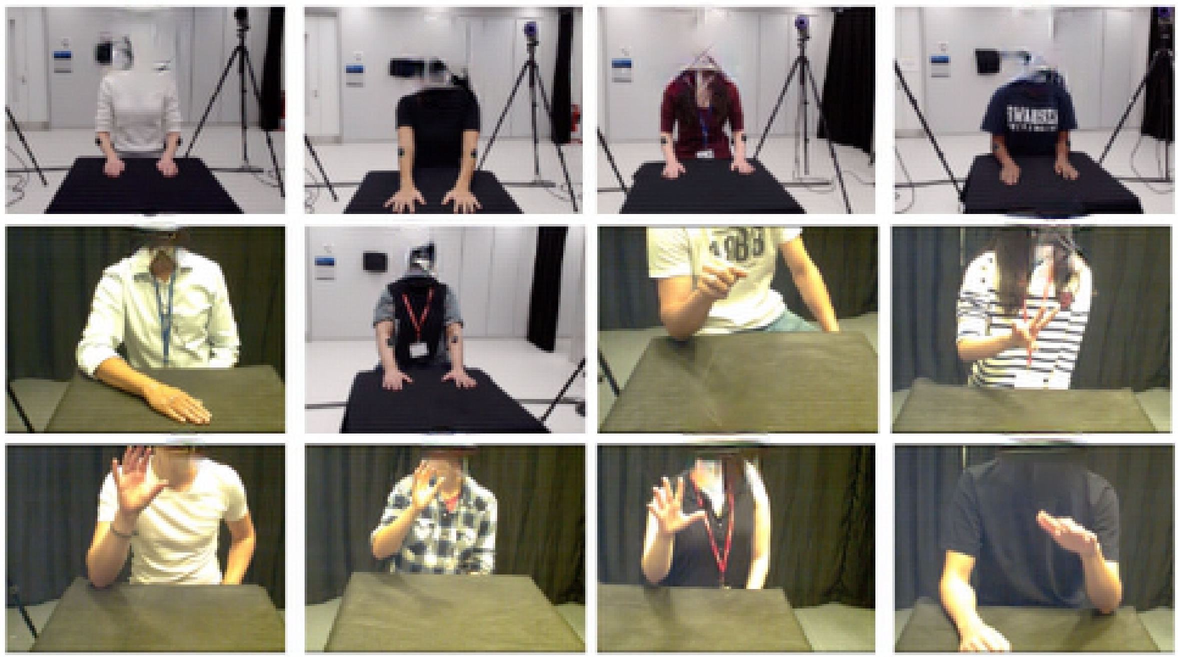
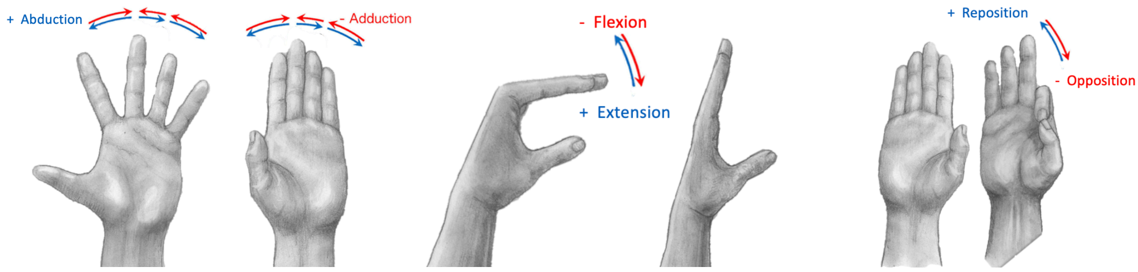


Figure 21: Twelve healthy participants at the Upper Limb Motion Analysis Laboratory at Imperial College London.

All participants involved in this investigation were healthy volunteers, with no hand impairment. Participants were asked to perform interventions relevant to improving ROM, selected from amongst the hand exercises previously adopted in literature (195,247,279).

The tasks performed in this study were chosen to include different numbers of degrees of freedom (DoF). The first exercise was finger abduction and adduction of the 2nd to 5th digits. During this task (Figure 22A), participants were asked to spread the fingers away from the long middle finger (abduction), and then to bring the fingers back, near the middle finger (adduction). This exercise was repeated four times for each hand. The second exercise was radial walking, which consisted of sliding the fingers one at a time towards the thumb. This task was repeated twice for each finger. The third exercise selected was metacarpophalangeal (MCP) joint flexion (Figure 22B). During the MCP flexion task, while sitting with forearm resting on table, participants were asked to bend the MCP joints of the 2nd to 5th digits. This activity was repeated twice for each hand. In the final task, participants were asked to perform thumb opposition (Figure 22C), placing the pad of the thumb opposite to the 2nd to 5th digits. They were instructed to bend the proximal interphalangeal (PIP) as much as possible. This activity was repeated twice for each hand. To summarize, a total of four hand exercises were performed including finger abduction and adduction, radial walking, MCP flexion, and thumb opposition.



(A) (B) (C)

Figure 22: Illustration showing three of the four hand exercises performed during the marker-based versus markerless investigation on healthy volunteers. The hand exercises include (A) abduction and adduction, (B) metacarpophalangeal flexion, and (C) thumb opposition.

After performing these exercises, the keypoints were extracted using both the marker-based and the markerless motion capture technologies, and the hand kinematics were measured and compared. The full experiment set-up is illustrated in Figure 23.

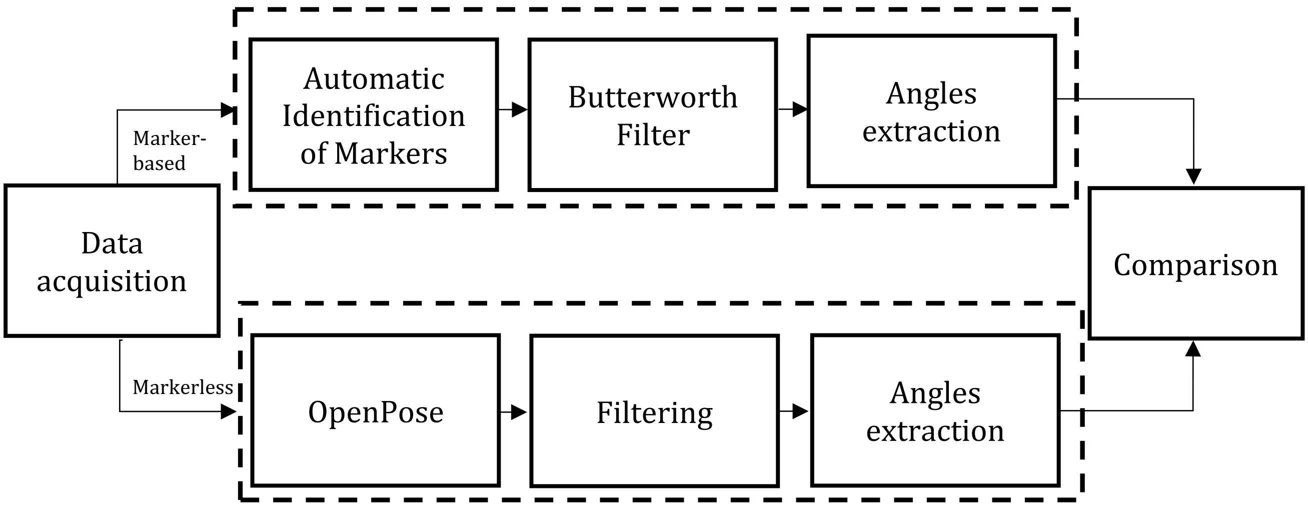


Figure 23: Flowchart of the experimental set-up of the marker-based versus markerless investigation.

4.2.2 Marker-based pre-processing

A total of twenty-six passive reflective hemisphere retro-reflective four millimetre markers were placed at specific positions on the dorsal surface of the right wrist, hand, fingers and thumb in accordance with the Hand & Wrist Kinematics (HAWK) (10) protocol. These semi-spherical markers were placed using double-sided adhesive tape, including the first, second, third, fourth and fifth proximal, intermediate, and distal phalanges. Markers were placed directly over the joint centres and on the fingertips on the distal border of the nail (Figure 24).

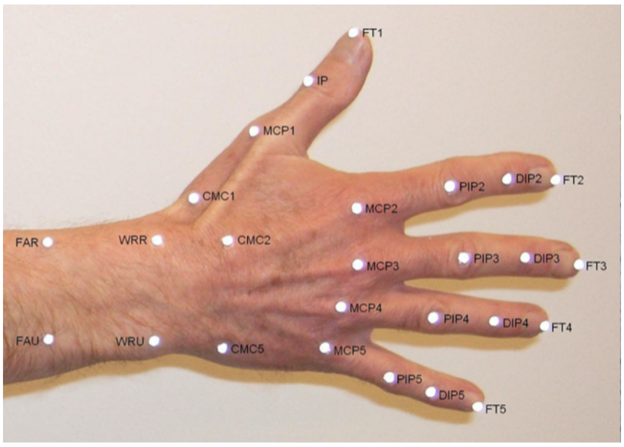


Figure 24: The Hand And Wrist Kinematics (HAWK) protocol with hemispherical marker placement. The image contains the marker placement for the distal heads of: the ulnar (WRU), the radial styloid (WRR), the 1st (MCP1), the 2nd (MCP2), the 3rd (MCP3), the 4th (MCP4), and the 5th (MCP5) metacarpals, the proximal phalanx of the thumb (IP), the proximal phalanx of the 2nd(PIP2), the 3rd (PIP3), the 4th (PIP4), and the 5th (PIP5) fingers, the medial phalanx of the 2nd(DIP2), the 3rd (DIP3), the 4th (DIP4), and the 5th (DIP5) fingers, the distal phalanx of the thumb (FT1), 2nd(FT2), the 3rd (FT3), the 4th (FT4), the 5th (FT5) fingers. The image contains the marker placement also for the dorsal aspects of the ulnar (FAU) and of the radius (FAR). The image contains the marker placement also for the proximal head of the 1st (CMC1), 2nd (CMC2), and 5th (CMC5) metacarpal at the carpometacarpal joint. Finally, the CMCVM visual marker is created halfway between the CMC2 and the CMC5. Reproduced with permission from (280).

The three-dimensional joint coordinates gathered from the markers were captured using an eight-camera Qualisys motion capture system (Oqus 500 + cameras, <0.4 mm error, Qualisys AB, Gothenburg, Sweden) and the Qualisys track manager (QTM) software. The data collection took place over two different acquisitions, with six participants in each acquisition. During the first acquisition, RGB video data were recorded using an Oqus RGB camera (Qualisys AB, Gothenburg, Sweden). Here, both the optical motion capture data and the video data were captured at a 30 Hz frame rate. In the second acquisition, the video data were recorded using an additional camera (Logitech StreamCam). During these recordings, the capture rate of the Qualisys system was set at 60 Hz and the Logitech camera was synchronized with the motion capture QTM software using custom-written LabVIEW (Laboratory Virtual Instrument Engineering Workbench from National Instruments Corporation) code. The QTM system was set to capture continuous recordings for 300 seconds for each hand, one hand at a time.

After the calibration and the synchronization of the system with the RGB external camera, participants were visually supported by a PowerPoint presentation that guided them throughout a set of hand tasks described above. These were performed with both the right and the left hands, while seated on a standard height chair with both feet flat on the floor.

4.2.3 Marker-based postprocessing

Several steps were carried out before extracting the joint angle computation, including labelling, refining, filtering, and segmenting the marker-based data.

Automatic Identification of Markers (AIM) is a function in QTM that automatically identifies and labels the trajectories tracked during a recording. Once a model is created, the connections between the markers are defined by the original model, with new trials that can be added to the model to give it additional examples of distances and angles between markers. Adding new trials to an AIM model will help the software apply it more easily to future test participants. Given this feature offered by QTM, a model was created in accordance with the Hand & Wrist Kinematics (HAWK) (10) marker placement.

Following the labelling, the smoothing tool in the trajectory editor of the QTM software was used to reduce spikes and noise in the data output from the motion capture system. A 2^{nd} order Butterworth filter with 5 Hz cut off frequency was selected due to the large frame ranges and presence of high-frequency noise. The filter served as a low-pass filter to attenuate information above the 5 Hz cut-off. Finally, the filtered data were manually segmented to isolate the different exercises for both the right and the left hands.

4.2.4 Markerless data pre-processing

Video data were captured using an Oqus RGB camera. The cameras were connected to the master computer that was utilized during the experiment and synchronized with the motion capture QTM software. The video data were captured from a frontal view of the participants, however the posture of the hand changed with respect to the camera throughout the trail. OpenPose (version 1.7.0) was installed from GitHub (CMU-Perceptual-Computing-Lab, 2020) and run with an NVIDIA Tesla K80 graphics processing unit under default settings to extract the keypoints. OpenPose, is a library written in C++ using OpenCV and Caffe that detects 21 keypoints on each of the hands.

To capture the hand ROM, the video data were first segmented into eight different exercises. Then OpenPose was executed on each frame of the video. The locations of twenty-one keypoints of participants' hands were independently estimated from each frame via OpenPose, as illustrated in Figure 25.

Figure 25: Keypoint visualization in output from OpenPose (17) that illustrates the inferred keypoints overlapped onto the image frames for healthy participants.

Data in output from OpenPose were visually observed. Instances where the fingers were incorrectly labelled due to the system swopping one finger with another, were manually labelled, assigning the correct value to the respective finger. Other inconsistencies, for instance, those where the fingers were incorrectly labelled and the tracking was missing due to intrinsic problems with OpenPose, were not manually corrected.

4.2.5 Markerless postprocessing

Once the finger keypoints were extracted using OpenPose a two-stage motion artefact filtering technique, previously implemented in similar studies using OpenPose on the lower-limb (18,20,177), was adopted to smooth the raw signal and decrease the noise generated by the architecture.

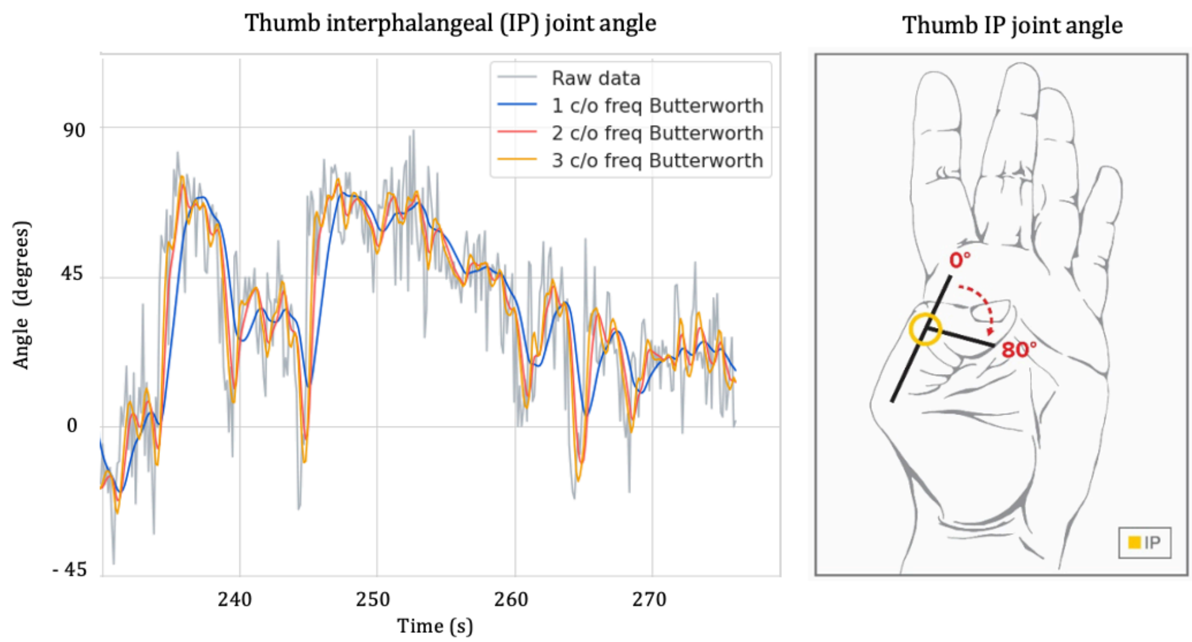
The first step of the two-stage motion artifice involved an outlier removal, the Hampel filter, which has been shown to be an accepted approach for outliers removal for raw signals series in output from OpenPose (281). The goal of the Hampel filter is to identify and replace the outliers in each series (282). The filtering technique removes the outliers by computing the median of a window comprised of current and adjacent samples and calculating the standard deviation of each sample using the median absolute deviation. If the considered sample varies from the window median by more than the threshold (dependent on the signal distribution) multiplied by the standard deviation, the filter replaces the sample with the median. The Hampel has two parameters to be tuned, the multiplying coefficient of the standard deviation (SD), that was kept at one and the window size that was set to four. No threshold was set for what was defined as an "outlier", opting for a visual inspection of the highest number of outliers identified an approach previously taken in the literature (283–285).

Following the application of the Hampel filter for removal of outliers, a generalized accepted approach to treating the raw signal in output from OpenPose is to smooth the raw signal further using a Butterworth filter (283,285). The Butterworth filter, is a one of the most used frequency filters to smooth raw kinematics data (286), and largely adopted to smooth raw signals in output from OpenPose (18,287), was defined as:

$$|H(\omega)| = \frac{1}{\sqrt{1 + (\omega)^{2n}}} \tag{4.1}$$

where ω is the cutoff frequency and n is the filter order.

A common challenge in the field of biomechanics is determining the cut-off frequency for the Butterworth filter (283,285). While some studies prefer the RANSAC approach, introduced in Chapter 2, the cut-off frequency based on the residual analysis proposed by Winter et al. (288) has been largely adopted for OpenPose signals suggesting a frequency selection below 3 Hz for a second order Butterworth filter (18,20). Furthermore, based on the cut-off determination using a residual analysis previous investigations reported that even with the forward and backward filtering (e.g. colloquially known as '*filtfilt()*') typically adopted to avoid phase lags, phase distortion was still observed at 1 Hz and 2 Hz (18,289). This was also observed in our results for cut-off frequencies at 1 Hz and 2 Hz (Figure 26). Therefore, based on benchmarking comparison, the second step of the two-stage motion artefact selected to smooth the signal was a zero-lag second order Butterworth filter with a 3 Hz cut-off. Therefore, the filtering technique was used as a low-pass filter to attenuate information above the 3 Hz cut-off frequency.



(A) (B)

Figure 26: (A) The effect of a zero-lag second order Butterworth filter with 1 Hz, 2 Hz and 3 Hz cut off frequencies (c/o freq.) applied to the OpenPose signal of the (B) thumb interphalangeal (IP) joint angle. Reproduced with permission from (343).

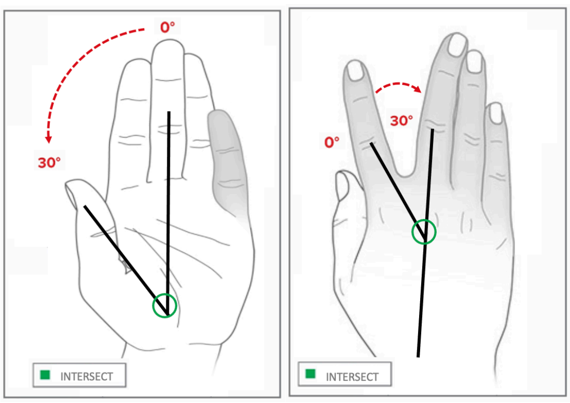
The consistency of the tracking system in inferring the two-dimensional keypoints from single frames showing hands with and without markers was investigated for one participant performing one action. The visible markers were removed first using CycleGAN (290) and then using an inpainting technique (291). More details on this reliability test are presented in Appendix A1.

4.2.6 Hand kinematics

Once the centres of the joints were located using both the maker-based and the markerless motion capture technologies, the hand kinematics were measured. DIP joints were considered to have one degree of freedom (DoF), PIP and thumb IP joints were considered to have one DoF, and MCP joints had two DoF. A total of 36 time-varying angular positions were measured for each participant, with 432 time series extracted for each methodology (marker-based and markerless).

The middle finger was used as a reference for the abduction and adduction task. The eight time-varying angles included the intersection between the thumb and the middle finger (Figure 27A), the index and the middle finger, the ring and the middle finger, and little finger and the middle finger, for the left and the right hands. Therefore, eight angles were measured for each participant during the abduction and adduction exercise.

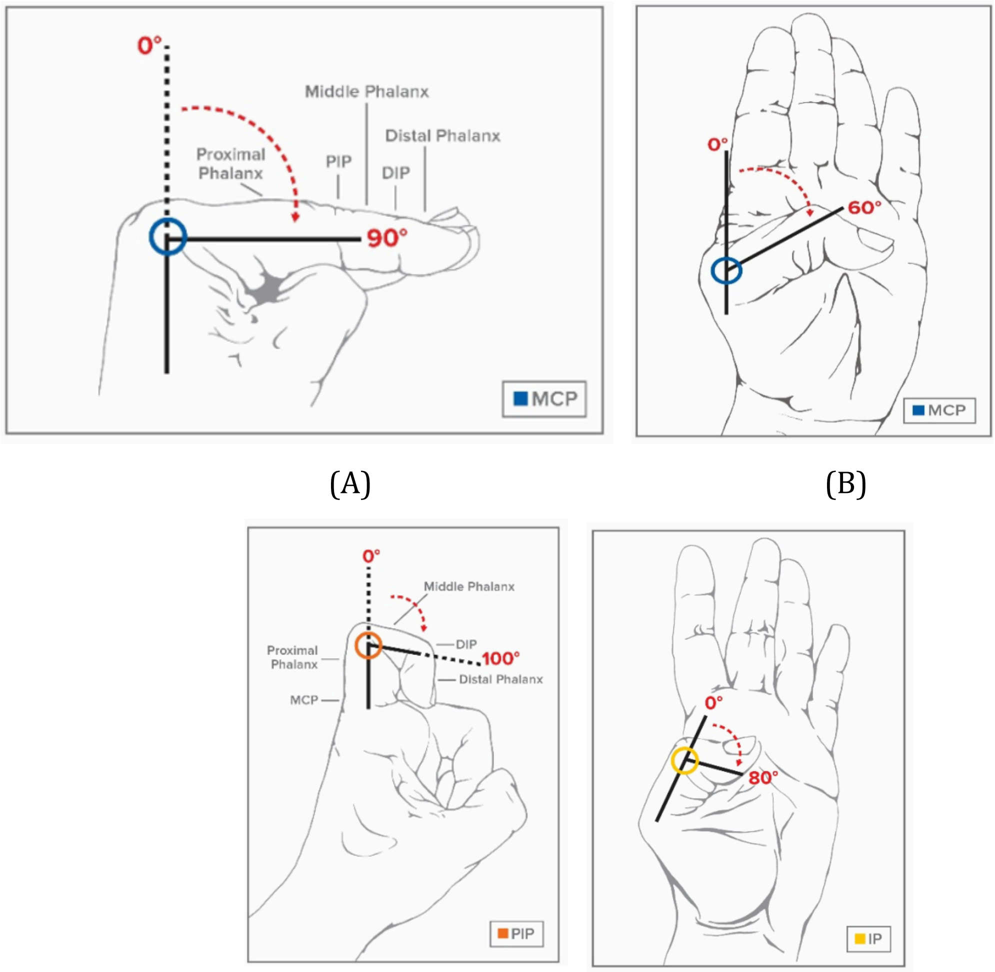
During the radial walking task, the reference digit was always considered the one that slid radially prior to digit performing the sliding. The eight angles measured included the intersects between the thumb and the index, the index and the middle (Figure 27B), the middle and the ring, and the ring and the little finger, both the right and the left hands.



(A) (B)

Figure 27: Measured position for the metacarpophalangeal (MCP) joint of the index finger (A), and of the thumb (B). Reproduced with permission from (343).

For the MCP flexion task the measured angles were the MCP angles of thumb, index (Figure 28A), middle, ring, and little fingers for a total of eight angle time series for the right and the left hands. Finally, during the thumb opposition, ten angles were measured. Those angles included the MCP of the thumb (Figure 28B), and IP of the thumb (Figure 28C), and PIP angles of the index (Figure 28D), the middle, the ring, and the little finger.



(A) (B)

(C) (D)

Figure 28: Measured position for the metacarpophalangeal (MCP) joint of the index finger (A), and of the thumb (B). Measured angles of the proximal interphalangeal (PIP) joint of the index finger (C), and of the thumb (D). Reproduced with permission from (343).

To describe the angles of the MCP, PIP and DIP joints, the included angles between the segments were determined. Using the segments illustrated in Figure 29, the angles were calculated as:

$$\alpha = arccos\frac{\vec{A}\cdot\vec{B}}{|\vec{A}|\cdot|\vec{B}|} \tag{4.1}$$

$$\beta = arccos\frac{\vec{B}\cdot\vec{C}}{|\vec{B}|\cdot|\vec{C}|} \tag{4.2}$$

$$\gamma = arccos\frac{\vec{C}\cdot\vec{D}}{|\vec{C}|\cdot|\vec{D}|} \tag{4. 3}$$

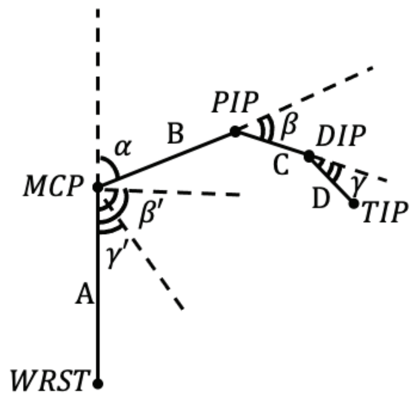


Figure 29: Illustration of a geometric representation of the finger. WRST is the wrist, MCP is the metacarpophalangeal joint, PIP is the proximal interphalangeal joint, and DIP is the distal Interphalangeal. α is the included angle of the MCP joint, β is the included angle of the PIP joint, γ is the included angle of the DIP joints.

Once the measurement error for $\vec{A}$, $\vec{B}$, $\vec{C}$, and $\vec{D}$ were defined as ε_n $n \in [A, \ldots, D]$, and a normal distribution of the error was assumed, the measured angles (^) with respect to the actual angles ($_0$) were:

$$\hat{\alpha} = \alpha_0 + \varepsilon_A + \varepsilon_B \quad \hat{\beta} = \beta_0 + \varepsilon_B + \varepsilon_C \quad \hat{\gamma} = \gamma_0 + \varepsilon_C + \varepsilon_D \tag{4.5}$$

Another measurement to assess hand kinematics is the Total Active Movement (TAM). The American Society for Surgery of the Hand (ASSH) defines the TAM to assess the ROM as the aggregate of the MCP, PIP, and DIP joints as:

$$TAM = \alpha + \beta + \gamma \quad \text{so} \tag{4.6}$$
$$\hat{S} = \hat{\alpha} + \hat{\beta} + \hat{\gamma} = \varepsilon_A + 2\varepsilon_B + 2\varepsilon_C + \varepsilon_D + \alpha_0 + \beta_0 + \gamma_0\varepsilon_D$$

The measurement of the TAM is error prone (50). Thus, the ASSH Total Active Flexion (TAF) is often used as a metric (292). Instead of adding the total measures captured with all the joints maximally flexed, the TAF refers to the measurement of active flexion of one digit (292). Thus, TAF isolates the maximum flexion angle minus the minimum flexion angle, for a given activity, for MCP, the PIP, and the DIP joints, as illustrated in Figure 30. Therefore, assessing the active flexion measures of joints under inspection for the specific exercise was selected as the preferred choice for this investigation.

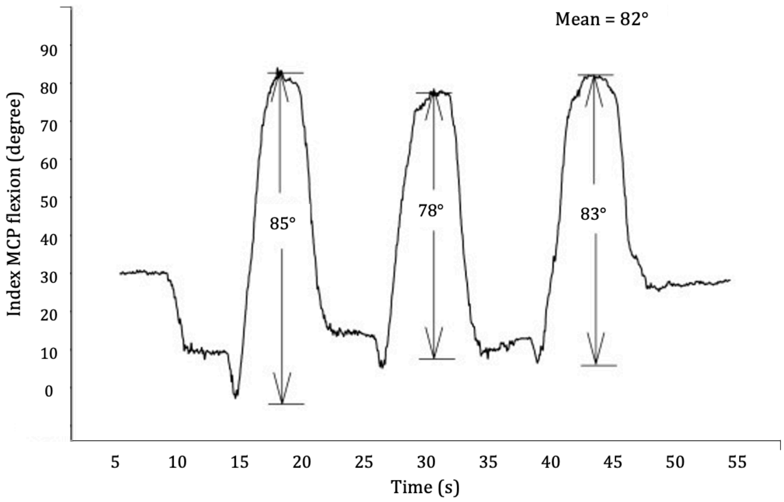


Figure 30: Trends indicating the total active flexion measurement for the index metacarpophalangeal flexion.

Once TAF was extracted for each digit and for each of the exercises under inspection, Bland–Altman plots (293) and linear regression were used to illustrate the agreement between the methodologies. In Bland–Altman analysis the agreement between two measures is assessed with the estimation of the standard deviation (SD) of differences with 95% limits of agreement (LoA) ±1.96 SDs of the mean.

4.3 Results

Representative plots for abduction and adduction (Figure 31A), radial walking (Figure 31B), MCP flexion (Figure 31C), and thumb opposition (Figure 31D) show the similarity between the two trends determined using OpenPose and obtained with the optoelectronic motion capture system, during the four tasks performed.

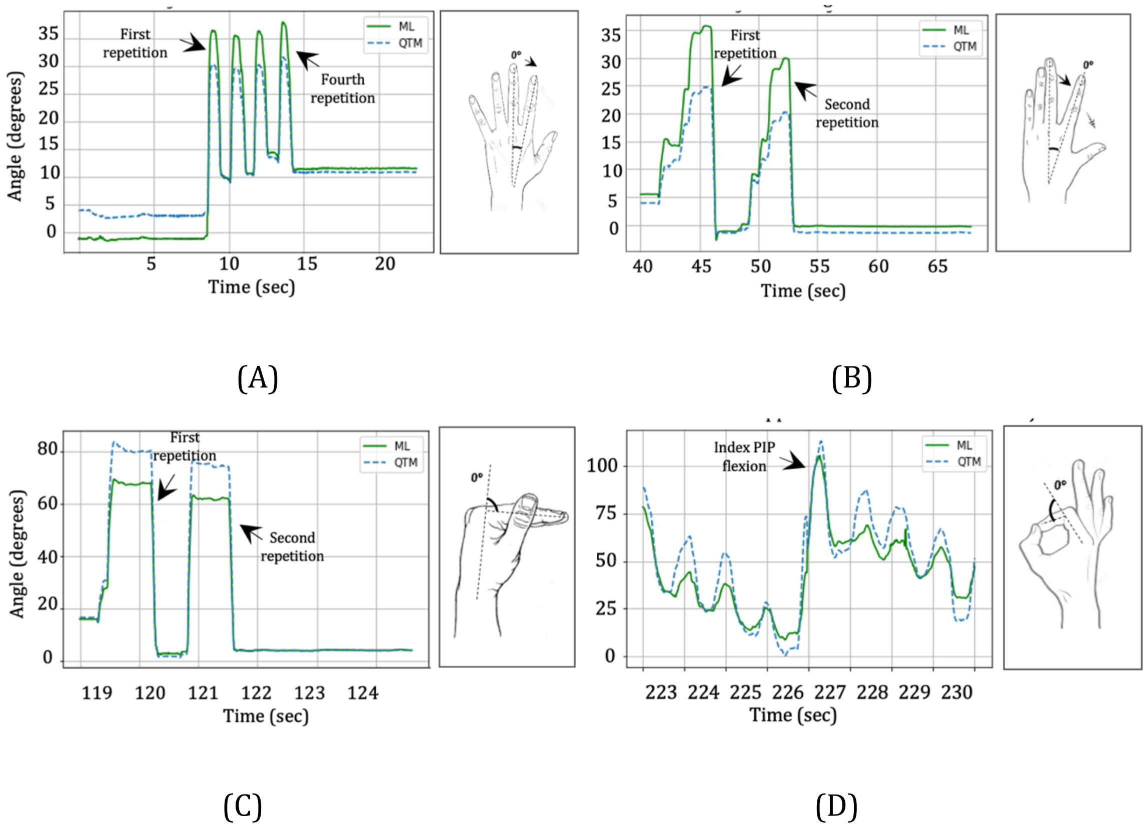


Figure 31: Examples of raw data for (A) index-to-middle finger angle for four repetitions of the abduction and adduction task, (B) index-to-middle finger angle for two repetitions of the radial walking task, (C) index metacarpophalangeal (MCP) joint angle for two repetitions of the MCP flexion task, (D) index proximal interphalangeal joint angle for the thumb opposition task, estimated using OpenPose (ML; solid lines) and measured with the optoelectronic system (QTM; dashed lines) for one representative healthy participant.

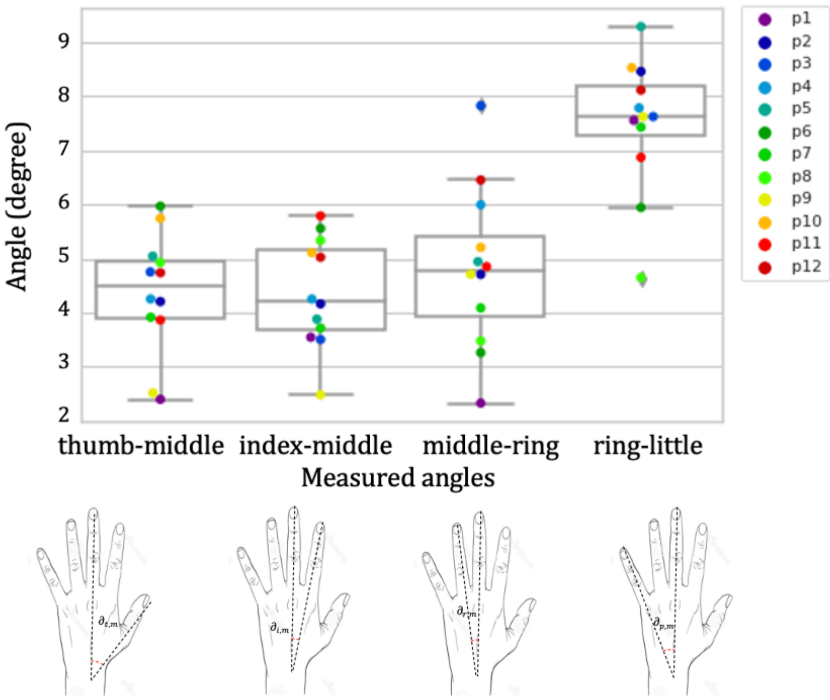
As a metric of comparison of the two-time series, once the angles were obtained from the two tracking techniques, the differences were computed using the root mean square error (RMSE) (Figure 32 and Figure 33). The predicted TAF measured for finger abduction and adduction, radial walking, MCP flexion, and thumb opposition was compared against the total TAF measured using the optoelectronic motion capture system using the Bland–Altman analysis and linear regression (Figure 34).

During the abduction and adduction activity, the finger kinematics estimated with OpenPose had a RMSE below 9° (Figure 32A), with the main error reported for the ring-to-little angle due to occlusion by the other fingers. Furthermore, the TAF values exhibited a mean difference between OpenPose and the optoelectronic motion capture system of 4.7° (Figure 34A) with LoA of 8.8° and 0.6°, and an R^2 of 0.73 showing good agreement between the two methods.

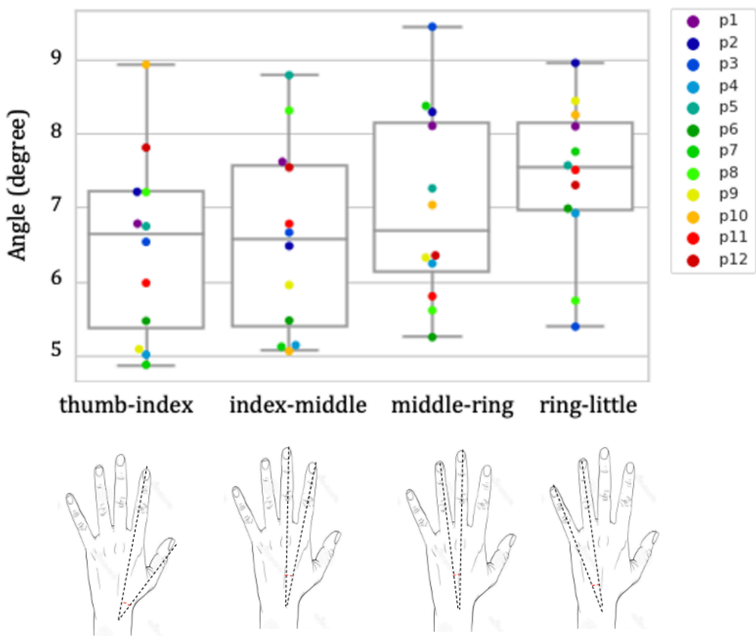
During the radial walking task performed on the table, the finger kinematics estimated with OpenPose had a RMSE below 9° (Figure 32B). The TAF values (Figure 34B) showed a mean difference between the methods of 5.0° with LoA ranging from 13.3° to -3.2° (Figure 6B). However, the coefficient of determination (R^2=0.40) suggested larger variability, compared to the abduction and adduction activity.

The MCP flexion exercise when comparing the two methodologies presented an error below 11° (Figure 33A), apart from two participants who had error values between 11° and 12°. The Bland-Altman plot (Figure 34C) presented a mean difference of 6.8° (Figure 6C) with LoA that go from 14.5° for the upper limit (+1.96 SD) to -0.8° for the lower limit. The comparison between the two methodologies yielded a modest R^2 value of 0.53.

Finally, during thumb opposition, the RMSEs (Figure 33B) were below 10° for 93.3% of the estimated values, while the other 6.7% reported an error between 12° and 14.5°. The main reason for the higher errors in 10% of values was occlusion by the other fingers, and OpenPose inadvertently swopping finger segment values. The mean difference between values (Figure 34D) was 4.7° with LoA 9.64° and -0.23°, and an R^2 value of 0.85.



(A)



(B)

Figure 32: Boxplots of root mean square differences between the OpenPose and the optoelectronic marker system during (A) finger abduction and adduction, (B) radial walking. Each colour represents a different subject.

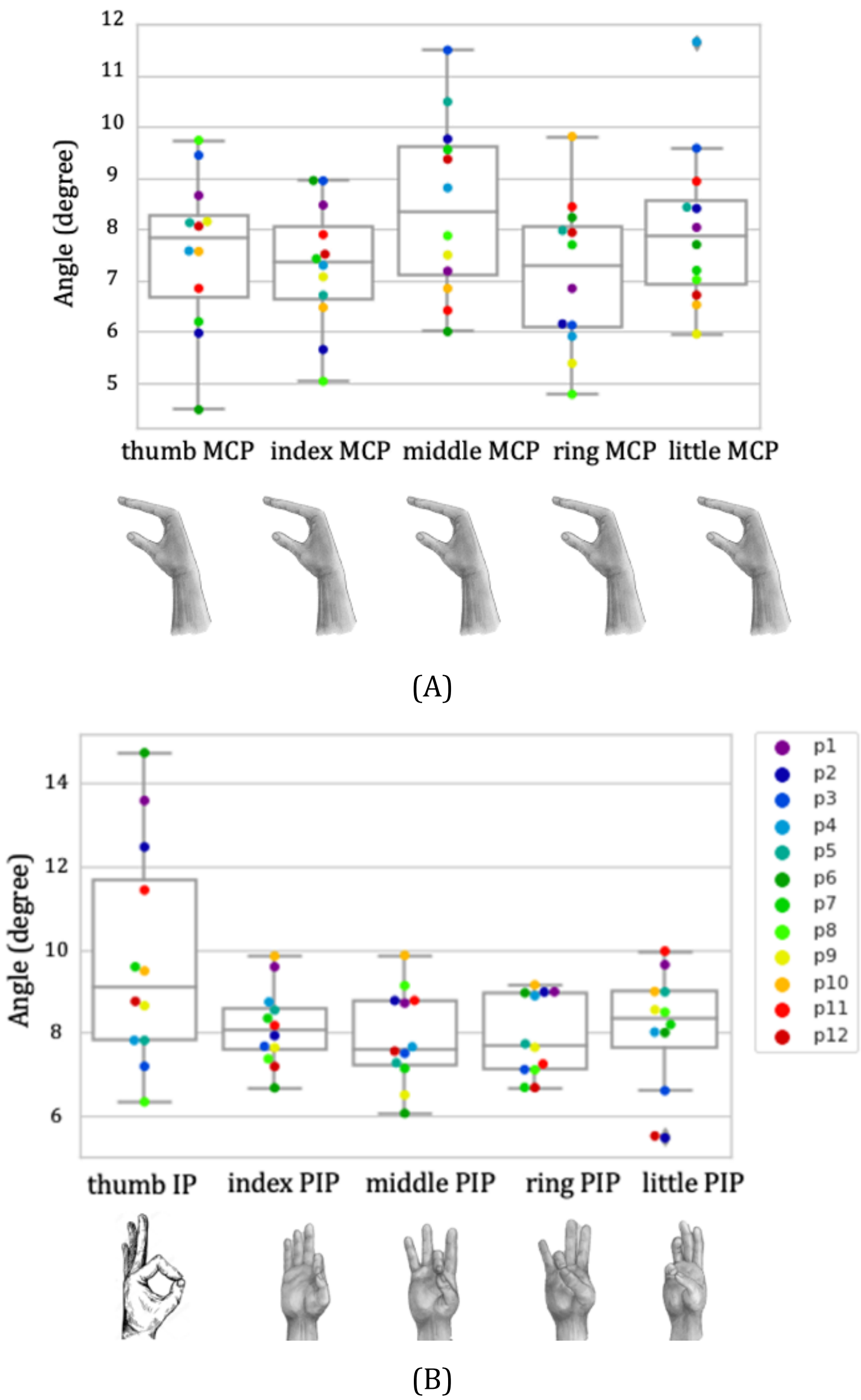


Figure 33: Boxplots of root mean square differences between the OpenPose and the optoelectronic marker system during (A) finger metacarpophalangeal (MCP) flexion and (B) thumb opposition. Each colour represents a different subject.

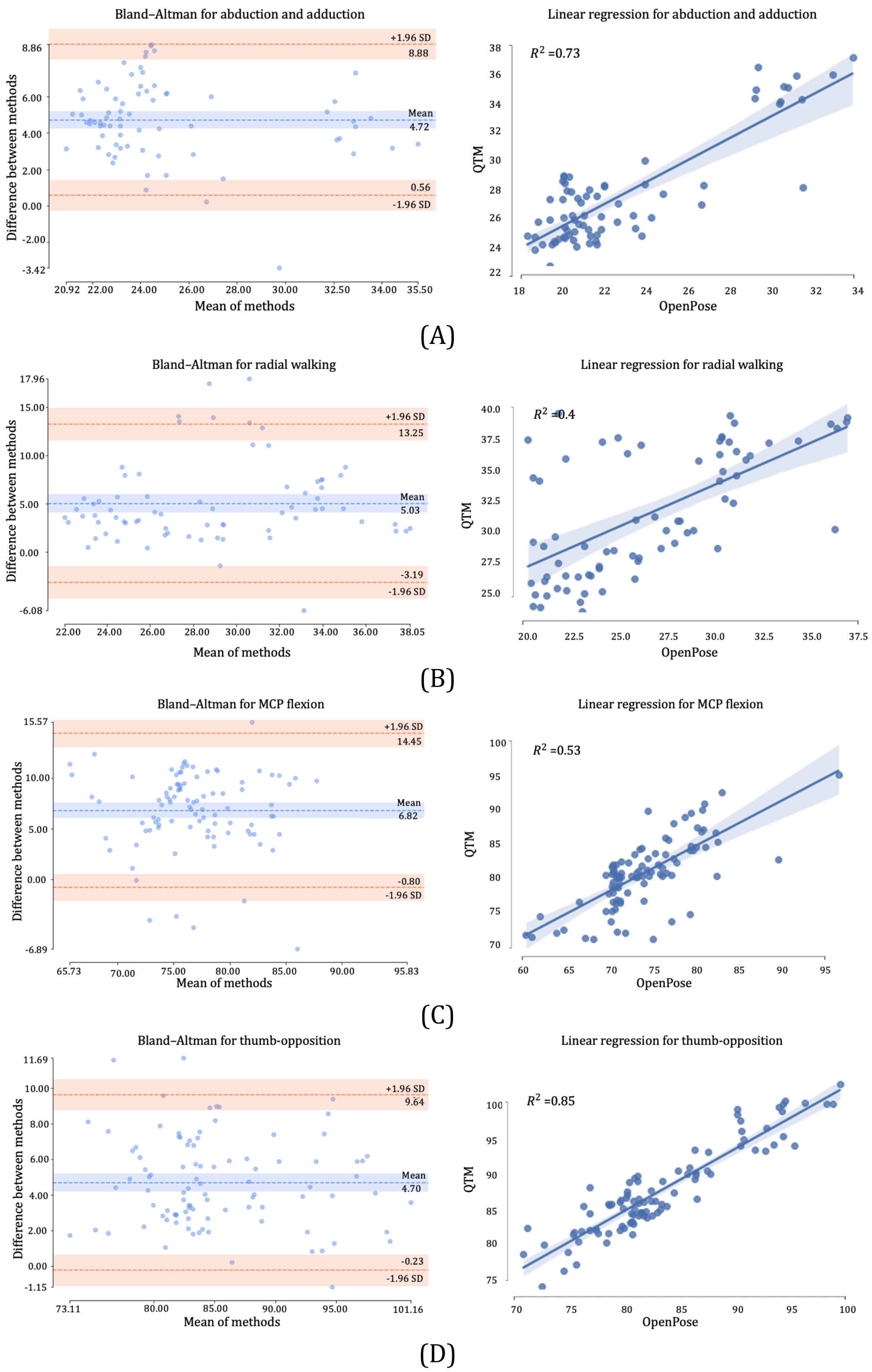


Figure 34: Bland-Altman plots (left) and linear regression (right) plots of total active flexion for (A) abduction and adduction, (B) radial walking, (C) MCP flexion and (D) thumb opposition of the 2nd, 3rd, 4th and 5th digits of the left and the right hands.

4.4 Discussion

This chapter proposes a tracking measurement system to assess specific hand activities using one monocular RGB camera. The technique uses a convolutional neural network (CNN), OpenPose, and two filtering techniques, the Hampel and the Butterworth filters, to evaluate a home-based method able to capture finger kinematics. The accuracy of OpenPose in tracking two-dimensional finger kinematics was assessed by comparing it with the three-dimensional finger kinematics obtained using a marker-based motion capture system.

Acceptable accuracy when comparing the two methodologies was defined based on instrument error of universal goniometers, which are trusted devices used in clinical practice to measure hand kinematics. Studies when comparing goniometer measurement against optoelectronic motion capture systems have reported an instrument error ranging from 2.4° to 9° in measuring finger kinematics (4,294). One drawback to using goniometry in clinical practice is that it involves lengthy processes and requires face-to-face assessments; thus, alternative solutions have been investigated over the past few years.

Research into markerless pose estimation has presented the potential for adopting commercial technologies to capture clinical hand metrics remotely. An example is the Leap Motion Controller™, which has been proposed as a portable and alternative solution to the gold-standard motion capture systems (128,295). However, most recently, Ganguly et al. reported an error that ranged from 19.31° to 28.29° when inferring the proximal and metacarpophalangeal joint positions using the Leap Motion Controller (272), making it unsuitable to replace gold-standard capture in clinical settings.

Markerless technologies that leverage deep-learning architectures have exhibited great potential for motion tracking, using monocular video cameras. For instance, two-dimensional pose estimation models have been validated for human gait (18,19,296), reporting an error of 5° to 15°. Leveraging these findings, this chapter offers a preliminary proof-of-concept investigation showing that hand pose estimation of hand kinematics using OpenPose can be reach similar levels of accuracy to assess kinematics during hand specific exercises. The comparison between the marker-based and the markerless technologies presented an error below 10°, apart with a few outliers that occurred with a 3.4% frequency rate.

Differences when comparing the two methodologies may be introduced by several factors, including the nature of the video recording. For instance, OpenPose depends on images labelled with keypoints, whereas marker placement relies on the physical location of anatomical landmarks. Another possible cause of misalignment that caused outliers could be linked to the comparison of the two-dimensional keypoints and the three-dimensional motion capture parameters. When calculating the included angle between two vectors from a projection of the three-dimensional landmarks onto a plane, fingers can still move in the three-dimensional space, leading to potential differences in the angle calculation. When assessing the other potential reasons for these outliers, self-occlusion was also observed.

Across the different hand exercises illustrated in this chapter, the coefficients of determination (R^2) presented good agreement (higher than 0.7) between the two methods for the abduction and adduction and the thumb opposition activities. Lower R^2 values, representing lower agreement between the two methods, were observed for the radial walking and the MCP flexion activities.

During the radial walking task, it was noted that the hand positioned vertically reduced the amount of keypoints lost, compared to when the hand was seated on the table. This was due to the nature in which OpenPose was trained to infer hand kinematics from monocular RGB cameras. Given the modest agreement of the two tracking systems during the radial walking task, and since the abduction adduction activity was able to extract the same joint ranges of motion as the radial walking exercise, it is noted that the abduction and adduction task would be the preferred activity for translation into clinical practice applications monitored using OpenPose.

The modest R^2 value (0.53) observed during the MCP flexion task can be attributed to the fact that during RGB video acquisition the 2nd, 3rd, and 4th digits were partially occluded by the 5th digit. Furthermore, it was visually observed that during occlusion OpenPose inverted the tracking, swapping the digits' values and causing visible errors for 18% of the dataset. This error could be mitigated by adopting visual manual postprocessing techniques, as proposed by Stenum et al. (19). However, this approach could not be automated and would limit the adoption of any activity into clinical practice.

OpenPose provides the joint centres locations together with the confidence values. When the confidence value was low, then error unrelated to occlusion, angle calculation, and the nature of the video recording was attributed to intrinsic parameters, as this tracking methodology does not estimate hand movements perfectly from frame-to-frame, as suggested in previous investigations (19).

To assess if visible markers applied on the participants' skin were introducing errors, two image-to-image translation techniques were tested to remove the remove visible marker's location. The techniques, known as CycleGAN (290)and image inpainting (291), are described in Appendix A1. As a result, it was noted that the presence of the markers did not compromise the performance of the markerless CNN tracking system.

The Bland-Altman plots (Figure 34) illustrated that the biases (mean differences) across the methods were consistent, ranging from 4.70° to 6.8°. Therefore, by offsetting the results with the consistent biases detected in these acquisitions, the accuracy of future results could potentially be improved. Given the constituency of the biases produced in output, further adoption of these findings would include an automated bias-correcting solution.

Despite the promising features demonstrated by pose estimation models to track fine movements of the human hands, video-annotation and manual segmentation still limits the scalability of this approach to clinical applications. An approach that would enable automated segmentation and video segment classification, leveraging video-level label data, could extend the capabilities of this investigation into clinical settings and provide the ability to examine larger volumes of video data in uncontrolled environments.

Chapter 5 Hand gesture recognition for automatic temporal segmentation

5.1 Introduction

Hand gestures are a vitally important form of non-verbal communication. The interpretation of hand gestures with wearable sensors (9,297), or cameras (162,298) aims to transform the gestures into meaningful instructions; this interaction is known as hand gesture recognition. The active field of hand gesture recognition has seen significant improvements over the past few years (155) and, most recently, combined with the latest advancements in computer vision, has encouraged the development of new technologies to support rehabilitation (159,299), robot control, and home automation (300).

Computer vision techniques rely on convolutional neural networks (CNNs) to extract two-dimensional (appearance-based) and three-dimensional (motion-based) array features. CNNs are generally used in image recognition to process pixel data. They take raw pixel data as input, train the designed architecture, and automatically extract features. These models have been divided into static (two-dimensional) and dynamic (three-dimensional) based on the model's output features. Several investigations (139,141,165) have implemented two-dimensional static appearance-based hand gesture recognition models (also known as two-dimensional CNN models) intending to develop a computationally inexpensive classifier to extract stable shapes of the human hand. However, these models do not consider the spatio-temporal parameters that occur from sequential frames of a video recording, as appearance alone cannot accurately identify the gesture signature (26). Therefore, new approaches, known as three-dimensional dynamic hand gesture recognition, have emerged to fill this gap.

Three-dimensional dynamic hand gesture recognition models also rely on CNNs, act like conventional two-dimensional CNNs, and have spatial-temporal filters. Since their introduction in 2015 (140), these models have been primarily embraced for hand gesture recognition (140,157,301), presenting excellent characteristics in recognising hand actions from both appearance and spatio-temporal features. However, they require more parameters than two-dimensional CNNs, meaning vast datasets, and making them more challenging to train (147). Furthermore, these approaches have additional drawbacks that include cost, the logistical challenges of dealing with complex and lengthy datasets, and the requisite quality of captured images needed for appropriate training. To overcome these drawbacks, previous research has leveraged a technique known as transfer learning (302).

Transfer learning is a methodology where architecture is implemented and trained on a specific activity and is then adopted for a different but linked activity (Figure 35). This technique is often employed to tackle the issue of a deficiency of training data (303,304). A usual objective of transfer learning techniques is to learn visual features from the initial assignment (304). This technique can train and acquire a forthcoming linked task from fewer data samples. Transfer learning is adopted when a novel, minor dataset is smaller than the dataset used to train the pre-trained architecture.

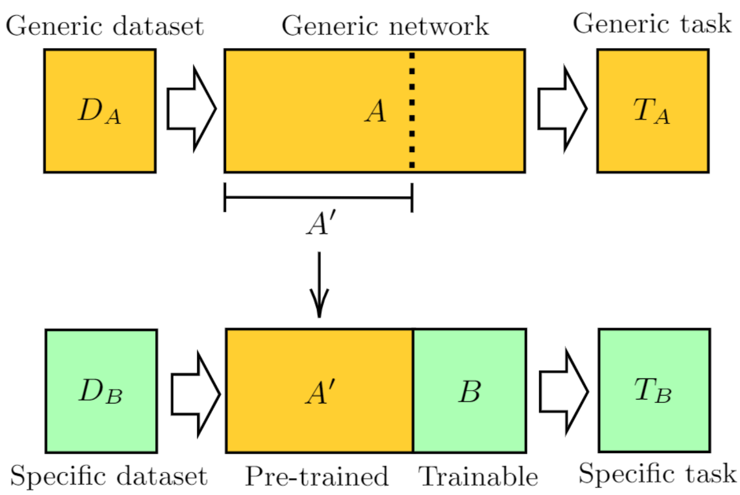


Figure 35: Schematic of the approach to transfer learning. Reproduced with permission from (305).

Another hurdle in dynamic gesture recognition for three-dimensional CNN represents recognising specific actions when dealing with continuous video streams (306). Identifying human activities within video sequences is difficult because of the vast irregularity of hand actions on a time scale, unclear frame quantity, distribution, and limits of gesture signatures. Furthermore, hand motions are often intricate and articulated and, when performed in an uncontrolled environment, can lead to occlusion that can limit the tracking. However, the ability to track and segment hand gestures in the real world can answer the need of applying these models to more realistic and generalisable tasks.

Manual segmentation of continuous video recordings is considered the most adopted technique when training hand gesture recognition (307). However, the process is lengthy, and often a large proportion of frames is left unlabelled, causing indexing issues in the training of novel classification methods. The ability to automatically detect action in video recordings has an essential function for different applications that require end-to-end process automation. But, while much work has been produced on increasing the accuracy of hand gesture recognition models and enhancing the strength of these approaches (138,155,298), just a few attempts have been presented for temporal segmentation (25,27).

Attempts at temporal segmentation have focused on motion trajectory (308) and skeletal tracking (309) from depth cameras. However, these systems were sensitive to the backgrounds and lighting conditions. A different approach, presented by Camgoz et al., suggested windowing the continuous video stream for segmentation (310). However, the length of the sliding volume was fixed, often cutting part of the critical features of the gestures. Moreover, appearance and hand motion information complement a temporal segmentation classifier (25). Still, Camgoz et al. also used only time-series data detected from hand motion, with no appearance information (310). In contrast, Wang presented a segmentation method that contained both action and appearance-based information, and used both RGB and depth capture modalities (25).

Increasing, enormous datasets of human movement are publicly available, as researchers seek to pool resources and work more openly. The 20BN Jester is a state-of-the-art dataset and the largest of human hand gestures collected from monocular RGB cameras.

It contains a total of 148,092 videos corresponding to 5,331,312 frames (157). Each video is, on average, three seconds, and the dataset contains a total of 27 classes.

This aim of this chapter is to present the training of a CNN using a small set of data and the development of a narrow architecture that can run efficiently for continuous hand gesture recognition. The key objectives of this chapter include:

1) To implement and to test the accuracy of a three-dimensional CNN model combined with a long-shot term memory (LSTM) unit to reliably classify and segment continuous video recordings and improve current manual-based segmentation when deploying models capable of executing tasks smoothly in real-world scenarios.

2) To evaluate the performance of transfer learning in implementing an architecture that is trained on a larger scale dataset, and then fine-tuned with a small-scale dataset.

3) To lay the foundations for a small-scale and reliable model, paving the way to a broader and optimised application that can be used to automatically detect where to run the keypoint hand tracking network.

5.2 Materials and methods

5.2.1 Experimental setup

Twelve healthy volunteers (six female, six male) participated in this experiment. All the participants were healthy, presenting with no hand pathology, no loss in mobility, and no experience of upper limb joint surgery or fracture in the six months preceding the data collection. All participants were also informed, both verbally and in writing, of their right to withdraw from the study at any time. Written informed consent was obtained from each participant. They were all able to speak and read English sufficiently to provide consent. The protocol was approved by the Imperial College Research Ethics Committee (ICREC). The entire pipeline adopted in the study is illustrated in Figure 36.

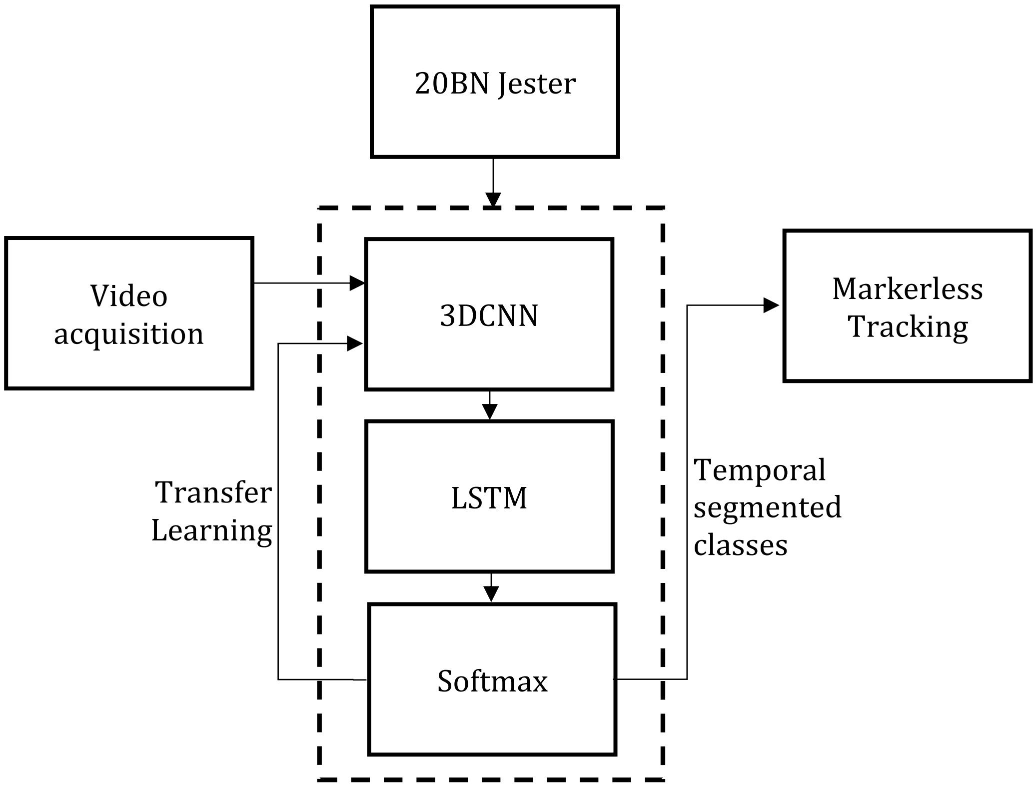


Figure 36: Flowchart of the experimental setup for the hand gesture recognition investigation. The pipeline uses transfer learning, pre-training the architecture on the 20BN Jester dataset (157), a three-dimensional convolutional neural network (3DCNN), a long short term memory (LSTM) and the output function (Softmax).

5.2.2 Data collection

Participants were asked to record one video sequence during online video meetings. To support the video data acquired by each participant, there was a timed PowerPoint to make the video acquisition consistent, to support participants on the activities to be performed during the recordings, and to inform participants on the way to position themselves relative to the device for the recordings.

To perform the hand gestures, participants were asked to use a standard device camera to capture the required hand exercises using any laptop, smartphone, desktop computer. A standard camera was defined as a camera developed from 2012 onwards that was able to capture video recordings at a rate of thirty frames per second. To assess if the data were captured from an acceptable browser and operating system, participants were asked to check that the specifications of the recording system were listed in Table 3.

Table 3: List of supported browsers and operating systems and browsers for online data acquisition using users' RGB cameras.

Type	Platform	Software Version Support
Mobile	Android	8 ('Android Oreo') and above
	IOS	Nine and above
Web	Chrome Desktop	Previous four major versions
	Chrome Mobile	Previous four major versions
	Firefox Desktop	Previous three major versions
	Firefox Mobile	Previous three major versions
	MS Edge	Current updated versions
	Internet Explorer	Current updated versions
	Safari Desktop	Previous three major versions
	Safari Mobile	Previous three major versions

Following the exercises presented in Chapter 4, the hand activities performed by participants in this part of the investigation included abduction and adduction, metacarpophalangeal (MCP) flexion, and thumb opposition. Each was performed four times with both the left and right hands. During these exercises, participants were asked to hold the position for five seconds. Four classes of gestures were defined based on the trials (Figure 37).

Figure 37: Illustration showing hand gestures classified during each trial: no gesture, abduction and adduction (Abd and add), metacarpophalangeal (MCP) flexion and thumb opposition.

The hand gesture sequences were captured from continuous video recordings of 250 seconds. The continuous video sequences were then manually segmented and labelled. Examples representing the data collected from these twelve participants are illustrated in Figure 38.

Figure 38: Examples of anonymized frames of the videos from the dataset of twelve participants. The images show the variance in the people's appearance and background scenes.

To improve the performance of the training and testing, the sample size of the video sequences was increased using the video data captured during the experiment described in Chapter 4. In Chapter 4, the video data were collected using a Logitech RGB camera from a total of twelve individuals. To ensure that the presence of visible markers on the hands did not influence the training of the gesture classifier, visible markers were removed using an image-to-image technique, known as image inpainting (291) (Figure 39). To avoid having the same healthy participant in the study twice, data from two participants were excluded. The acquired dataset consisted of video sequences collected from 22 participants performing three different activities, intercut by "no gestures".

(A) (B)

Figure 39: Comparison of (A) original images with visible markers on the hand and (B) image inpainting (291) technique with markers removed.

In addition to the captured data, the 20BN Jester dataset acquired by Materzynska et al. (157) (Figure 40) was used. The classes of interest in this study, including "no gesture", "abduction and adduction", "MCP flexion", and "thumb opposition", were not present in the Jester dataset. These specific hand activities, also validated in Chapter 4 to markerlessly extract kinematics, were considered relevant for their adoption in clinical settings.

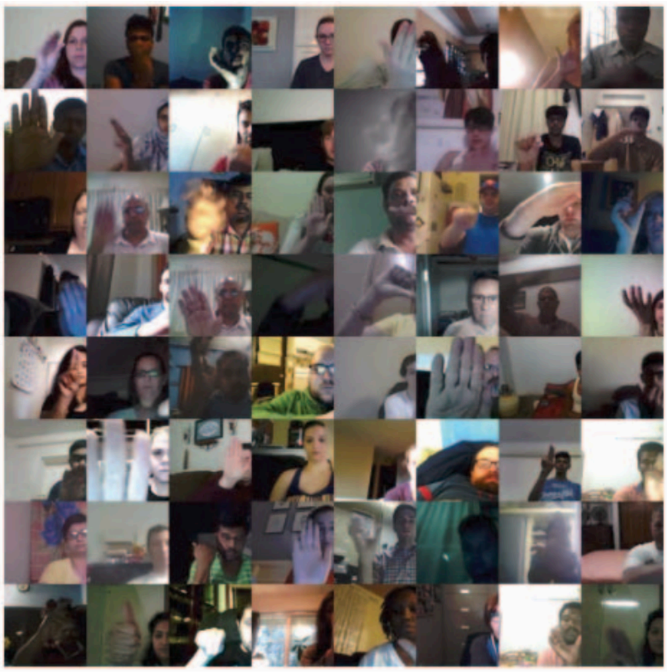


Figure 40: Examples of videos from the public 20BN Jester dataset. Reproduced with permission from (157).

5.2.3 Pre-processing

The normalisation of the captured frames was necessary to ensure that each input to the three-dimensional CNN had the same distribution, ensuring each class had the same number of frames. This was particularly important as, although the timing of the participants' actions was marked by the PowerPoint presentation, individuals could execute a hand gesture at different speeds. Ideally, a three-dimensional CNN input should always be balanced, making the model converge faster. When the input frames are not normalised, the weights could have different calibrations across features, making the cost function converge ineffectively.

To address concerns about normalisation, the frame length was equal for all the acquisitions for which the hand gestures were at the centre of the video (139,146). Following the structure of the 20BN Jester dataset, the normalisation imposed a fixed length, set to be 32 frames. If the number of frames was higher or lower, a down-sampling or a padding function was applied, respectively, to generate fixed-length videos. Given the S_n sequence of RGB frames, the L_S length of the sequence, and the L_F fixed length, the padding and down sampling techniques were defined as:

$$S_n = \begin{cases} padding(S_n), & L_s < L_F \\ (S_n), & L_s = L_F \\ downsampling(S_n), & L_s > L_F \end{cases} \tag{5.1}$$

Following normalisation, the images were resampled to be 64 × 64 pixels, given the need to expedite classification. The labels were assigned manually, and the videos were manually trimmed for input into the segmentation classifier. Finally, for training and validation, the dataset was split into training, validation, and testing sets, with a 70:20:10 ratio.

A total of 2,812 short video sequences of healthy volunteers performing three different hand activities, each separated by "no gesture", were collected, of which 1,968 short video sequences (≈ 70% of the dataset) for training and 845 data points (≈ 30% of the dataset) for validation and testing. Each short video sequence contained 32 frames for 89,984 frames in total. When the additional video data captured at the Upper Limb Biomechanics Lab were added, a total of 5,155 short video sequences were collected, of which 3,609 short video sequences (≈ 70% of the dataset) for training and 1,546 data points (≈ 30% of the dataset) for validation and testing. Each short video sequence contained 32 frames for a total of 113,410 frames for training and 6,784 for validation.

5.2.4 Postprocessing

After the data pre-processing, the architecture was implemented based on an existing model originally introduced by Tran et al. (311), known as C3D. Specifically, a modified version of the C3D network, similar to the multimodal RGB-D-based network by Hakim et al. (26), was considered. Furthermore, to make sure that the three-dimensional CNN model was able to learn longer sequences, another unit, able to acquire long-term temporal features, was combined with the three-dimensional CNN, an LSTM unit. The final architecture (Figure 41) consisted of a three-dimensional CNN layer with three convolutional layers, a Rectified Linear Unit (RELU) as activation function in the hidden layers used to avoid vanishing gradient, one LSTM layer, a flatten layer, a fully connected dense layer and an activation function, also known as the Softmax layer.

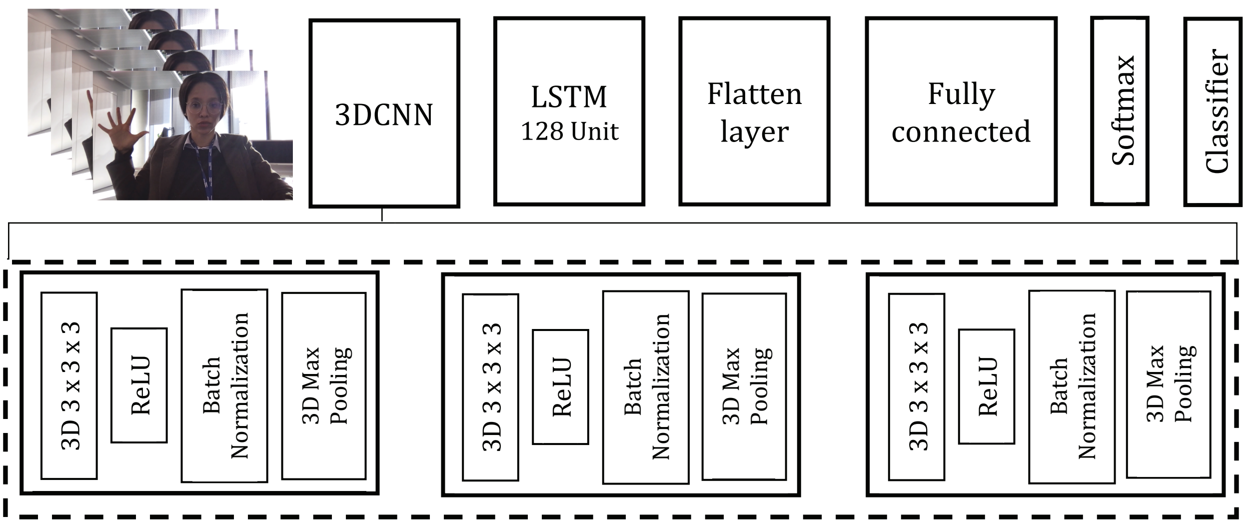


Figure 41: Three-dimensional convolutional neural network (3DCNN) with long short-term memory (LSTM) for dynamic hand gesture recognition. The video sequence is fed into the 3DCNN to operate 1-D & 3-D convolutions for time & space dimensions. The hidden layers (dashed box) with Rectified Linear Units (ReLU) as activation function the 3DCNN has limitations to learning long-term information and therefore, the vector goes

into an LSTM. The tensor in output is then flattened into a single dimension, inputted into a fully connected layer and finally, the activation function (Softmax) predicts the classes. The multi-dimensional input tensors were flattened into a single dimension. The flattened layer is often employed in the presence of multi-dimensional output. This layer aims to produce a linear output that can be conveyed onto a dense layer. A dense layer (also called fully connected) joined every input neuron to every output neuron in the preceding layer. Finally, the Softmax function produced a vector that denoted the list of probability classes of possible results. Based on the output from the Softmax the frames were then segmented into those where the activities occurred and those where there was no gesture. The class "no gesture" was provided in case no activity was performed, but also for frames without a hand, when participants placed the hand down following a performed activity.

The CNN model was trained in Google Colaboratory and the TensorFlow framework was used to deploy the model (312). The baseline model was pre-trained on five classes of the 20BN Jester dataset, including count-to-five, swiping down and left, thumb-up, and thumb-down. These activities were selected to include different image frames of isolated digits and the palm with all the digits for both the left and right hands. Starting from the architecture trained on the above mentioned five classes belonging to the 20BN Jester dataset, a technique known as transfer-learning (303) was then used to fine-tune the model to the activities performed in this study. The technique took the parameters from the previously trained model, froze the last layers to avoid the weights in the last (frozen) layers being updated, and then new trainable layers were added, together with new data to fine-tune the model.

A total of four tests were performed. During the first two tests, transfer learning was used with three convolutional layers. Then, to increase performance, an additional convolutional layer and an increased sample size were considered. The first two tests were evaluated over mini-batches of 13 epochs, following the segmentation classifier proposed by Wang (25). The last two tests were evaluated over a batch size of 64 epochs, training batch also presented in Wang (25) investigation. The 12GB NVIDIA Tesla K80 graphics processing unit provided by Google Colaboratory was used for training of the 20BN Jester dataset for the baseline model, which took approximately nine and a half hours. For the first and the second tests, the training times were respectively one and a half hours and two and a half hours, while for the last two tests, they were two and four hours.

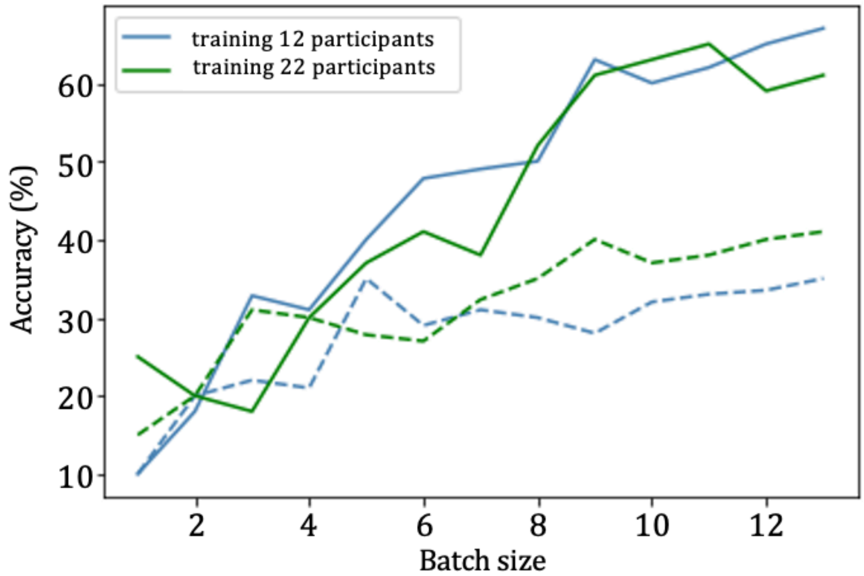
The classifier accuracy was determined using the function '*numpy.argmax*' to obtain the highest predicted class scores for each data point. Furthermore, the metric used to evaluate the performances of the model was the Jaccard index or intersection over union value (313). The index is often used for segmentation classifiers and was computed to analogise a set of predicted labels with a set of the corresponding true labels. Letting A and B be the set of frames predicted and ground truth manually labelled, respectively, the index is defined as:

$$JACCARD = \frac{|A \cap B|}{|A \cup B|} \tag{5.2}$$

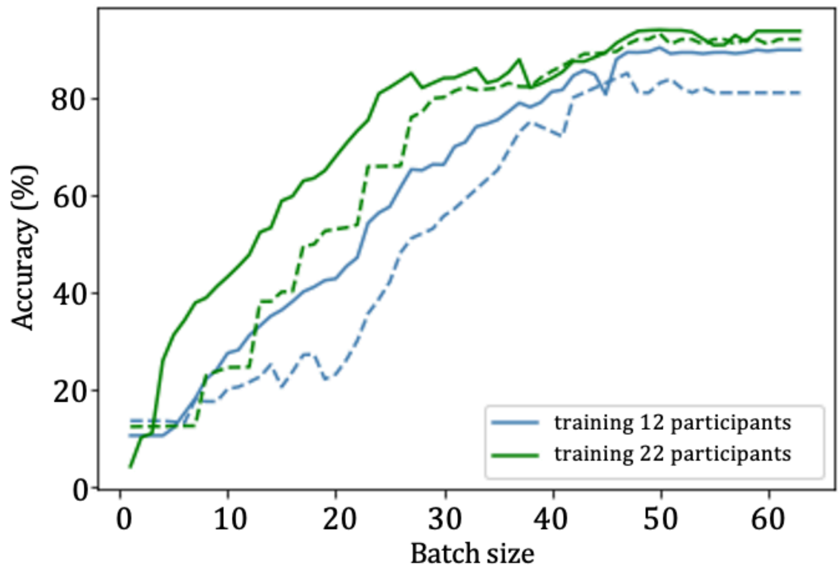
The Jaccard index varies from zero to one, the larger is the index, the higher is the accuracy of the segmentation classifier.

5.3 Results

Training and validation accuracies for 13 and 64 epochs for 12 and 22 participants show limited levels of accuracy (below 70%) reached for 13 epochs (Figure 42A) and increased level of accuracy (93.95%) reached for 64 epochs (Figure 42B). In the training and validation curves illustrated for 64 epochs, the training performed on 22 participants outperforms the training on 12 participants. Overfitting was observed during training after 50 epochs, in both cases (12 and 22 participants), suggesting that additional training would not result in the model having improved learning.



(A)



(B)

Figure 42: Results of the training and validation for a training batch of (A) 13 epochs (batch size) for 12 and 22 participants and (B) 64 epochs for 12 and 22 participants. Dashed lines indicate validation curves.

A representative output from the Softmax function (Figure 43) of the temporal segmentation for a continuous video recording for the three-dimensional CNN hand gesture classifier trained for 64 epochs and 22 participants illustrates the agreement with manual segmentation (ground truth).

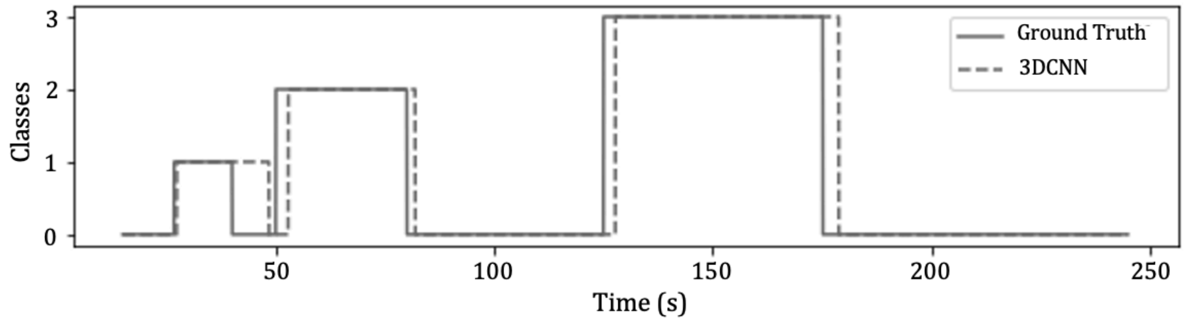


Figure 43: An example of the temporal segmentation and classification in output from Softmax function of the three-dimensional convolutional neural network for 64 epochs and 22 participants (dashed lines) compared against the ground truth manually segmented for Participant 1 for the labels "no gesture" (class=0), "abduction and adduction" (class=1), "metacarpophalangeal (MCP) flexion" (class=2), and "thumb opposition"(class=3).

The mean Jaccard index and the accuracy percentage values for 64 epochs for 12 and 22 participants for all the recordings are illustrated in Table 4.

Table 4: Comparison of the three-dimensional convolutional neural network for 12 and 22 participants using the mean Jaccard index $\overline{J_s}$ and the accuracy percentage (%).

Dataset	Number of frames	Mean Jaccard Index $\overline{J_s}$	Accuracy (%)
12 participants	89,984	0.794	83%
22 participants	113,410	0.812	93.95%

The training runs, executed for batch-sized 64, computed an initial mean Jaccard index that reached 0.794 (± 0.44), increasing to 0.812 (± 0.105) for the enlarged sample size of 22 participants. The validation accuracy showed 83% (± 0.05), increasing to an accuracy level of 93.95% (± 0.37) when additional participants were included. The "no gesture" label agreed with the manually segmented ground truth 96.47% of the time for all participants. The "abduction and adduction" class agreed with the ground truth 92.5% of the time for all participants. The "MCP flexion" label agreed with the manually obtained labels 95.7% of the time for all participants. Finally, the "thumb opposition" class was in agreement with the ground truth 90.93% of the time for all participants.

5.4 Discussion

This chapter illustrates a CNN that automatically classifies and segments a single video containing specific hand exercises including no gesture, abduction and adduction, MCP flexion, and thumb opposition. The segmentation of continuous video recordings was based upon a classifier that identified when the label "no gesture" was present. The presented pipeline addressed the challenge of hand gesture recognition from long video sequences captured using a monocular RGB camera.

The implementation of the three-dimensional CNN was based on a model known as C3D, proposed by Tran et al. (311) and made of an high-resolution and a low-resolution sub-architecture, both trained individually. Even if the C3D model presented good performance, the cost of training two different models is high, so a modified version, which incorporated the two networks into one, was used in this chapter. This modified C3D, however, could only detect short temporal characteristics from short video sequences, while the aim of this chapter was to introduce a network that detects short-term temporal features from long video sequences. Therefore, the final CNN was combined with an LSTM unit, capable of learning the long-term dependencies in long video sequences.

The studies previously presented that combined three-dimensional CNN with LSTM units for hand activity recognition used both RGB and depth modalities to extract the motion signature (25,26), while the three-dimensional architecture implemented in this chapter was only based on an RGB sequence, showing a similar level of accuracy (93.95%) can be reached also from a single acquisition modality. Furthermore, the proposed network outperformed the 82% accuracy presented by Hakim et al. (26). The overfitting observed after 64 epochs was similar to that of other investigations that used dual modalities (25, 26). The use of transfer learning to reach an acceptable (above 80%) level of accuracy enables the possibility of scaling this approach to include different hand gesture activities, showing how the model can be trained effectively on a small dataset to create an effective small-size segmentation classifier.

The mean Jaccard recorded in Wang's study was 0.6127 for the RGB modality (25), while in this investigation the mean Jaccard reached 0.794, outperforming the value presented in Wang's investigations. However, Wang's accuracy was based on the Montalbano hand dataset, containing different hand activities from those implemented in this investigation. Therefore, further investigations would be needed to compare the performances of this network using this metric. Furthermore, no inconsistency was shown across the segmented video recordings for action and participants, meaning that segmentation accuracy was not based on specific actions or on specific participants.

The end-goal for this model is to enable a more effective markerless inference of two-dimensional keypoints extracted from frames of video recordings when the relevant hand gestures are recorded. The markerless keypoints detection architecture validated in Chapter 4, OpenPose (17), is a much larger and more computationally expensive model compared to the three-dimensional CNN introduced in this chapter. This approach would enable users to upload video data and obtain hand kinematics from OpenPose run only on the relevant sub-sets of frames, instead of the entire video length, creating a more computationally efficient approach with an inexpensive computational classifier that is more accessible and adaptable to mobile applications in a way that overcomes capacity restriction.

To adopt and scale this application in real-work scenarios, if multiple classes are considered, future directions could include testing this approach for real-time application using a finite state machine system that can decrease the classes under inspection and increase the accuracy for real-time application. To further improve the model's performance for real-time applications, the input image size or the number of layers could be increased. On top of the 20BN Jester dataset, an additional dataset could be used to enhance the model's performance. The Jester dataset was developed by actors and did not provide numerous occlusion cases. Regardless, in realistic circumstances, occlusion exists. Recordings captured in unconstrained scenarios may incorporate additional types of interference, such as blurry hand gestures if the participants or the camera moves suddenly during the acquisition. Rescuing identifiable cues of image interference for a real-time hand recognition model would be an attractive research direction.

Furthermore, while the supervised-based transfer learning produced expected outcomes, the approach presented in this chapter could be transported to unsupervised learning and could support the automated labelling and segmentation of long video recordings, increasing the models' generalizability.

Most recently, there has been the diffusion of novel architectures for deep learning, such as the "transformers" that aim to replace more traditional recurrent modules, such LSTMs, for long sequence interactions (314). These techniques have been explored to a limited extent, with one study that uses them for dynamic hand gesture recognition (315). However, that study included depth modality to reach a level of accuracy similar to the one presented in this investigation, for longer sequences. Furthermore, these novel transformer techniques are computationally more demanding, compared to LSTM techniques (316). Therefore, LSTM was the final choice when trading off long-term precision with computational complexity.

Adapting current gesture recognition techniques to the specific mobility short exercise sequences would have benefits that go beyond this single application. A real-time device that requires minimal manual processing could process and identify multiple gestures as soon as an image frame is received. This approach could be deployed into online hand gesture recognition studies for advanced assistance systems, surveillance, aided robotics, and clinical applications. For instance, the pipeline illustrated here could be integrated into remote monitoring clinical solutions, presenting the training of a model that uses a smaller dataset implemented on a small architecture that can run efficiently to solve the classification problem for hand temporal segmentation. This would pave the way to a broader application in hand tracking models, incorporating other hand activities categories, and obtaining a more generalizable approach, that would include different hand exercise programmes and different hand conditions.

Chapter 6 Clinical proof of concept: a non-contact measure to monitor hand movements in rheumatoid arthritis

6.1 Introduction

Rheumatoid arthritis (RA) is an autoimmune condition causing soreness due to swollen and damaged joints. It affects around 1% of the world's population and treatments depend upon drugs that repress the immune system, adjusted according to the disease severity. Clinical trials over the last three decades have established that the best clinical outcomes are achieved when RA participants are treated-to-a-target (Figure 44) of low disease activity (310–312). To accomplish this, RA patients demand frequent monitoring. The best outcomes have been observed in studies that reviewed disease activity on a bi-monthly-basis, titrating medication accordingly (310–312).

In 1990, a landmark change in RA management was the introduction of objective measures to quantify swollen joints and effectively assess disease activity using a technique known as the Disease Activity Score (28 Joints) (DAS-28) (48). The DAS-28 is nowadays considered in clinical practice the principal system to objectively measure disease activity changes in RA. The system includes inspection of joints within the hands, wrists, elbows, shoulders, and knees.

The work presented in this chapter was partially funded by the 'Proof of Concept Award Competition for Translational Musculoskeletal Technology Projects', programme within the Bioengineering Department at Imperial College London supported by the Wellcome Trust Musculoskeletal Medical Engineering Accelerator.

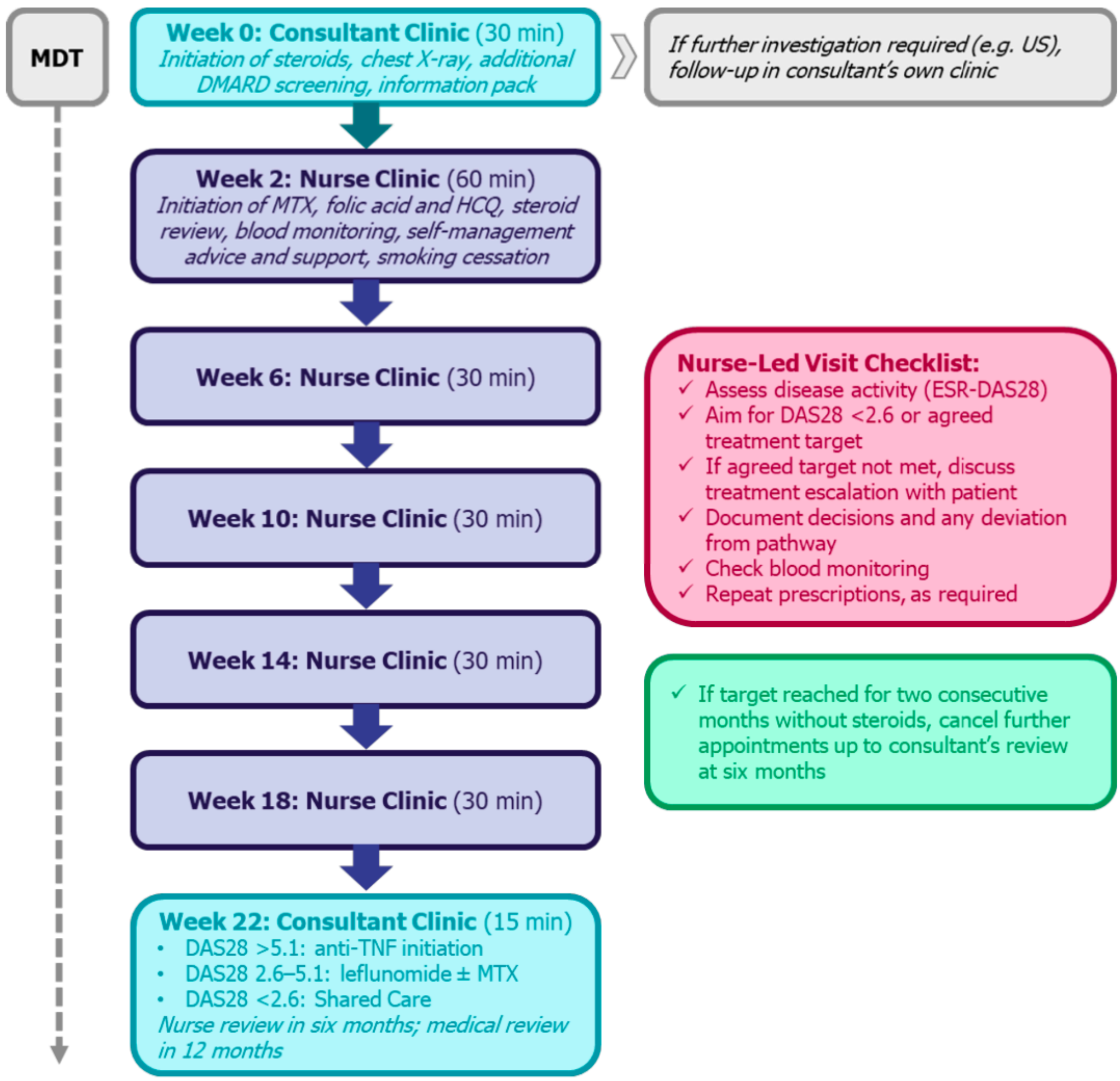


Figure 44: Rheumatoid arthritis (RA) treat-to-target scheme. Early inflammatory arthritis patients can join the disease-modifying anti-rheumatic drug (DMARD) titration pathway even if the three-week window from their general practitioner referral to first appointment has been missed. MDT: multi-disciplinary team; US: ultrasound; MTX: methotrexate; HCQ: hydroxychloroquine; DAS-28: Disease Activity Score (28 Joints); Reproduced with permission from The British Society for Rheumatology (320).

The DAS-28 involves four domains: clinician reporting swollen joint count (SJC), clinician reporting tender joint count (TJC), a global measure of pain using a visual analogue 100-mm long horizontal scale, and a biomarker of inflammation from a blood test, including the C-reactive protein (CRP) or erythrocyte sedimentation rate (ESR) levels. The different items are inserted into a calculator that provides in output a score between zero to nine, with scores below 2.6 reflecting excellent disease control (or patient being in remission), while scores above 5.1 indicate very active disease. The DAS-28 measurement takes around 30 minutes to record, and currently requires RA patients physically present in the clinic.

Many studies have highlighted the necessity for a practical alternative that can capture and quantify disease activity remotely (48,321,322). In the past two years of the COVID-19 pandemic, there has been an obligatory move towards virtual consultations, taking place using video meetings (323), with the option to remotely prescribe medications based on the visible signs of illness progression. The service of video consultations can offer visual clues that previously only could be acquired during the in-person assessment. However, it is conceivable to miss valuable clinical information with virtual assessments, as such methodologies can lack objectivity. The telehealth virtual consultation is based upon clinician observation and qualitative assessments but could be theoretically improved using digital home monitoring devices.

Remote monitoring offers the possibility of determining disease activity remotely. Most systems tend to isolate single components of the DAS-28 and use these to estimate disease activity (319). For instance, the biomarker levels has been obtained using home commercially available ESR/CRP testing kits (324). However, CRP and ESR are not condition-specific tests and can only observe if an inflammation is ongoing, which makes them not reliable as the only metric to track disease activity in RA (325). Other studies have illustrated the possibility of capturing qualitative patient pain scores (226). Nevertheless, such qualitative assessment can be unreliable and not condition specific. For TJC, several investigations have illustrated a good correlation between joint tenderness learnt through patient self-assessment and the scored measured by clinicians (244,326). However, the same studies have found only a very poor correlation in patient self-assessment for SJC.

The limitation in capturing SJC in RA has paused the integration of these remote monitoring tools into RA clinical practice, including virtual consultations. The main reason for this is that joint swelling has been considered the most critical component of the DAS-28 as: i) it can cause a reduced range of motion (ROM) (327), ii) it is a key predictor of future damage within a joint, iii) its absence can point towards alternative diagnoses such as fibromyalgia (328). Crucially, medications for RA do not help fibromyalgia, and if RA therapies were erroneously prescribed for fibromyalgia, the drugs could result in harm for patients (328). The assessment of swollen joints is aggravated in the small joints of the hand, where SJC could serve as an early indicator of disease activity, thus allowing timely adaptation of treatment plan procedures (224). It is, therefore, vital that the SJC is accurately assessed.

Given the importance that swelling plays in the small joints of the hand for RA patients, several hand interventions have been implemented to improve hand mobility and restore swollen joints (329). These exercise programmes are performed using participants' own cameras and represent a low-cost and widely accepted remote procedure, currently part of the NICE pathways to manage RA (35). These interventions have demonstrated increased ROM and improved swollen joints in the short term (330). However, they do not provide any feedback on the improvements each time the intervention is performed, are not linked to the clinical records of the patients and, for these reasons, studies have demonstrated that they lose adherence (256,331).

To emphasize the importance that small joints of the hand play in RA disease activity, in 2010 the American College of Rheumatology and the European League Against Rheumatism (ACR/EULAR) implemented a joint distribution criteria (Table 5) to classify disease activity based on the involvement of small joints of the hand (320).

Table 5: Disease activity quantification based on the swollen joints of the hand according to the 2010 American College of Rheumatology/ European League Against Rheumatism (ACR/EULAR) Classification Criteria for Rheumatoid Arthritis (RA).

RA joint involvement	Disease activity
<1 swollen joint	remission
1-3 swollen joints	low disease activity
4-10 swollen joints	active disease activity
>10 swollen joints	very active disease activity

Objective quantification of the ROM in small joints of the hand for RA patients has been investigated in several studies (40,265,268,332,333). The aim of these investigations was to implement a mechanism to record and assess how many joints are swollen, defining ROM thresholds for flexion and abduction under which joints could be classified as being swollen. The ultimate goal, once the SJC has been implemented, was to determine the disease activity status for RA patients based on swelling in the joints of the hand. These studies have reported a 17-28% reduction in ROM when the condition is at its early stages, increasing to 35-49% after eight years (265). Based on these findings, thresholds under which these joints could be classified as swollen in relation to the ROM deficits for individuals who have had RA for more than eight years are summarised in Table 2.

These investigations have been implemented using different techniques, ranging from gold-standard optoelectronic motion capture to goniometers. Due to financial and time constraints that deter the use of gold-standard optoelectronic motion capture, goniometric assessment has been preferred in the clinic, and is performed by rheumatology nurses. However, these assessments require face-to-face consultation, which is not always feasible given the periodicity at which the condition needs to be monitored. A system that would provide objective tracking of hand kinematics could also infer disease activity and would be able to support more frequent monitoring and improve treat-to-target approaches.

The latest advancements in deep neural networks together with the ubiquity of standard video cameras provide the opportunity to gather clinical endpoints remotely. Amongst other convolutional neural network (CNN) approaches discussed in Chapter 3, OpenPose (17) has been shown to obtain reliable two-dimensional (x, y) keypoints from monocular RGB cameras without the requirement for markers or special gloves. This tool has been validated against gold-standard optoelectronic motion capture systems for gait kinematics and has been adopted on clinical populations, e.g., cerebral palsy (28) and Parkinson's disease (20).

Leveraging these validations against optoelectronic systems for the CNN-based two-dimensional (*x, y*) keypoint detectors, these algorithms have been applied to video data gathered during telehealth consultations. Rosique et al. (334) implemented telerehabilitation software based on two-dimensional keypoints extracted using OpenPose for the lower limb. Similarly, Chua et al. (335) implemented a telehealth system that used video captured during remote consultation to infer posture and angular movement for patients with mobility impairments in the lower limb.

In Chapter 4, the feasibility of applying OpenPose to infer the centres of the joints of the digits, including metacarpophalangeal (MCP), proximal interphalangeal (PIP), and distal interphalangeal (DIP) joints, has been validated for specific hand exercises against an optoelectronic motion capture system. Therefore, similarly to the above-mentioned studies validated for the lower limb and applied to telehealth consultations, this chapter embraces this validation to deliver a pipeline that can be used to deliver objectivity to remote assessments.

One potential challenge of this approach is that identifying the motions of interest with videos provided by patients. Being able to break down the inputs of OpenPose into specific segments according to motion type makes its application more scalable and relevant to remote motion assessment. Several CNNs incorporating long short-term memory (LSTM) have been described in Chapter 3 to address this task. In Chapter 5, a classifier, trained on an available hand gesture dataset, 20BN Jester dataset (157), was fine-tuned using transfer learning to the gestures of relevance in this study. These hand exercises are often adopted innervations to restore loss in mobility and reduce swelling in rheumatoid hands. The segmentation technique described in Chapter 5 could be used to partition the videos and reduce the computational costs, so that OpenPose outputs can be considered only when the relevant actions are taking place.

The first aim of this chapter is to validate a remote monitoring tool that uses specific hand exercises for individuals with RA to determine disease status based on swollen joints count (Figure 46), by considering the following objectives:

i. To collect anonymized RGB video recordings of individuals with RA performing hand activities, gathered with the disease activity score measured in the clinic.
ii. To determine the accuracy of temporal segmentation of continuous video recordings into sections containing pre-specified hand exercises.

iii. To compare the inferred disease activities and the number of swollen joints disease activity observed during the virtual consultation, both compared against the ground truth.

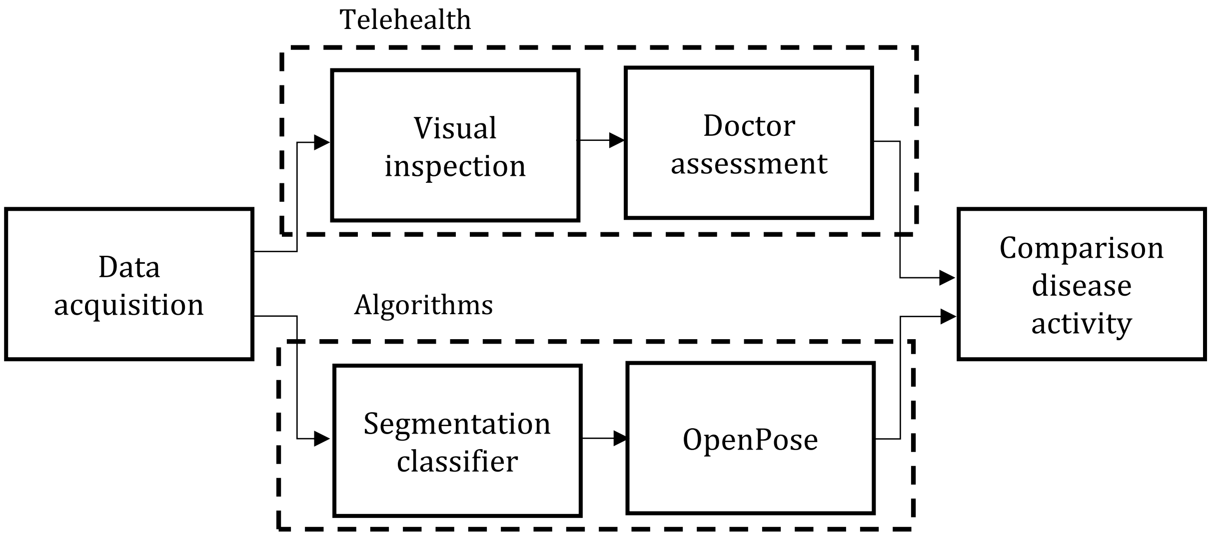


Figure 45: Flowchart of the pipeline illustrating the telehealth and the algorithm-based approach to assess disease status. Both are compared with the ground truth gathered in the clinic.

The second aim is to indicate if the presented pipeline can provide a means to navigate through the new territories of remote virtual consultations, enabling more frequent disease activity monitoring that exploits RA patients' own technologies.

6.2 Materials and methods

6.2.1 Participants and data collection

The protocol was approved by the King's College Research Ethics Committee and included permission for the Imperial College London principal investigators and co-investigators to have access to the data and utilize the data for researcher purposes. The inclusion criteria involved collecting videos of adults over 16 years of age, willing to anonymously take part in the study, with a diagnosis of RA that met the 2010 ACR/EULAR criteria (37) for more than eight years, and with regular hand anatomy (defined as no missing fingers). A requirement of the study was for participants to be able to record an

RGB video from any device, with the hand in the field of view of the camera, not occluded by clothing, with good lighting conditions.

Participants with RA were identified, and video recordings collected with the support of the National Rheumatoid Arthritis Society (NRAS), a patient-led association in the UK specialized in RA and juvenile idiopathic arthritis. The NRAS's letter of support is in Appendix A2. To maintain consistency in the video acquisition, the timed PowerPoint presentation adopted in the Chapters 4 and 5 was shared with NRAS and provided to participants for the video recordings. Written informed consent was obtained from each participant. Eleven RA patients (eight female, three male) participated in this study Figure 46.

Figure 46: Cropped sections of videos of the ten participants with rheumatoid arthritis performing hand function activities captured using a monocular RGB camera.

Individuals with RA participating in the study reported a median disease duration of eleven years, ranging from eight to thirty-five. The video acquisition was set for 250 seconds each for a total of 72,500 frames collected. Each recording showed RA patients performing hand exercises including abduction and adduction, metacarpophalangeal (MCP) flexion, and thumb opposition. The exercises were repeated twice for each participant with both the left and the right hands. Alongside each video, disease activity scores were collected in person, within the five days previous to the video recording, looking at all the four components of the DAS-28 including global symptom measure, CRP blood test, SJC, and TJC. The video recordings of the eleven RA patients performing hand activities and the corresponding disease activity associated to each participant were electronically shared by the charity.

Once the video recordings were obtained, a virtual consultation was simulated with an experienced rheumatologist consultant at King's College Hospital NHS Foundation Trust. The rheumatologist was not previously instructed on the ground truth of the video recordings and did not receive any information on the participants prior to the consultation. The simulated virtual consultation consisted of one asynchronous telehealth video call. Asynchronous telehealth, also known as “store-and-forward”, has been adopted in clinical practice during COVID-19 to simulate synchronous consultations (taking place for all parties involved at the same time) during remote assessments (336). To simulate an outpatient appointment taking place by video, each pre-recorded video was presented to the rheumatologist who observed the patients doing the exercises to label the joints with mobility compromised and suggested a classification of low, active, and very active disease activity. This was done to deliver a visually based impression of the presence of swollen joints.

6.2.2 Data processing

All the videos received electronically from the NRAS were manually anonymized (faces blurred) before executing the segmentation classifier. Of the eleven video recordings received, one recording from one participant was found to be corrupted; therefore, it could not be used in this investigation. All the video recordings were acquired at a standard resolution of 640 × 480 and observed to ensure that the hands were always present in the field of view of the camera, and that the timing was consistent.

Before applying the segmentation classifier, the recordings were manually labelled. Video sequences were classified as belonging to one of the four classes, including "no gesture", "abduction and adduction", "MCP flexion", and "thumb opposition". The continuous video sequences were then temporally segmented to only consider the relevant parts of the videos where the exercises occurred in the three-dimensional CNN classifier with the long-short term memory (LSTM) unit for temporal segmentation illustrated in Chapter 5 to isolate the activities were the hand exercises occurred. The code was executed locally, and the results from the network were compared against manual segmentation to evaluate the accuracy of the technique on an impaired population. The classifier performances were assessed using the accuracy percentage (%) and the Jaccard index $\overline{J_s}$ (313). The Jaccard index varies from zero to one. The larger is the index, the higher is the accuracy of the segmentation classifier.

For each frame of the segmented video sequences, only the regions where abduction and adduction, MCP flexion, and thumb opposition exercises took place were considered, while the transition frames where no exercise occurred were eliminated. OpenPose was used within a Google Colaboratory notebook. The OpenPose output delivered: 1) a JavaScript Object Notation file for every video frame containing pixel coordinates (origin at the upper left corner of the video) of each keypoint detected in the frame, reported points were the estimated (x, y) coordinates, in pixels, of the centres of the MCP, proximal interphalangeal (PIP), distal interphalangeal (DIP) joints and tips of the 2nd, 3rd, 4th and 5th digits, for the right and left hands; 2) a new video file with a stick figure overlayed on top of the original video recording that represents the detected keypoints (Figure 47).

Figure 47: Visualization output from OpenPose output that illustrates the inferred keypoints overlapped onto the image frames for four individuals with rheumatoid arthritis.

The raw data were then filtered. Following the approach in Chapter 4, the multiplying coefficient of the standard deviation (SD) of the Hampel filter was set to one, and the window size was set to four. A Butterworth filter with a cut-off frequency of 3 Hz was also applied. The combination of the two filtering techniques is a widely adopted approach to smooth the sequence in output from OpenPose (283,284).

Once the raw signals were obtained and filtered, the time-varying angles were extracted for each exercise. From the abduction and adduction activity the four time-varying angles corresponding to the finger intersect, defined as the intersection between each finger with the middle finger, were extracted for each hand. From the MCP flexion exercise the MCP angles were estimated. Finally, for the thumb opposition activity the PIP time-varying angles were inferred for both the right and the left hand.

Once the time-varying angles were extracted for each frame, the total active flexion (TAF) angles for each exercise were determined. TAF separates the highest flexion angle and the lowest flexion angle for a given hand exercise. The average values of the TAF were estimated, and the number of swollen joints were assessed, based on the value in Table 2. Finally, based on the number of swollen joints the ACR/EULAR criteria (Table 5) were adopted to classify the disease status.

Given the biases provided in output in Chapter 4, bias-correcting was evaluated. The TAF values to estimate the number of swollen joints were unbiased by offsetting the output with the bias values observed in the Bland-Altman plots. The bias values were 4.72° for the abduction and adduction activity, 6.82° for the MCP flexion task, and 4.70° for the thumb opposition task.

Finally, to evaluate the level of agreement between the two methodologies, the Cohen's kappa correlation coefficient (337) was determined. The usage of this coefficient was based on studies adopting it in the literature when comparing ratings against scoring indexing systems (28, 345). This metric assessed the agreement between the two assessments performed by two raters (Rater 1 and Rater 2). The Cohen's kappa coefficient can be expressed as a contingency table (Table 6), that expresses the distribution of the variables given two assessments and two raters, where x_1 represents the total number of instances that both raters were in agreement, x_2 indicates the total number of instances that caused disagreement between the two raters due to Rater 2 being incorrect and Rater 1 correct, x_3 indicates disagreement between the two raters representing the total number of instances that Rater 1 was incorrect and Rater 2 correct, finally x_4 represents the total number of instances where both raters were incorrect. The Cohen's kappa correlation coefficient is calculated as:

$$\kappa = \frac{P_0 - P_e}{1 - P_e} \tag{6.1}$$

where P_0 indicates the detected agreement probability derived by adding the number of times the raters agreed (x_1+ x_4) and dividing it by the total number of tests (x_1+ x_2+ x_3+ x_4), and P_e refers to the predicted agreement probability, obtained by the total number of instances Rater 1 rated "Correct" (x_1+ x_2) divided by the total number of tests, multiplied by the total number of instances Rater 2 rated "Correct" (x_1+ x_3) divided by the total number of tests, added to the total number of times that Rater 1 rated "Incorrect" (x_3+ x_4) divided by the total number of tests, multiplied by the total number of times that Rater 2 rated incorrect (x_2+ x_4) divided by the total number of tests.

Table 6: An example of a 2x2 grid to interpret results of the Cohen's kappa correlation.

		Rater2	
		Correct	Incorrect
Rater 1	Correct	x_1	x_2
	Incorrect	x_3	x_4

6.3 Results

From the assessment in the clinic, one RA participant was classified with very active disease activity (DAS-28 score over 5.1), two were categorised with active disease activity (ranging from 3.2 to 5), two more were classified with low disease activity (ranging from 2.6 to 3.2), and six were in remission (below 2.6). One participant in remission had the video corrupted and was the only one not used in the study. The disease stages provided by the rheumatologist, who visually observed the video data during the simulated virtual consultation determined that eight participants were in remission and two participants had active disease activity (for one of the two the disease activity was classified as being very active). These results reported by rheumatologist during visual inspection and the results of the approach proposed in this study are illustrated in Figure 48, both are compared against the ground truth measured in the clinic. The figure illustrates the disease activity level obtained from the pipeline implemented in this chapter has a higher agreement with the ground truth compared with the assessment obtained through an asynchronous telehealth video consultation.

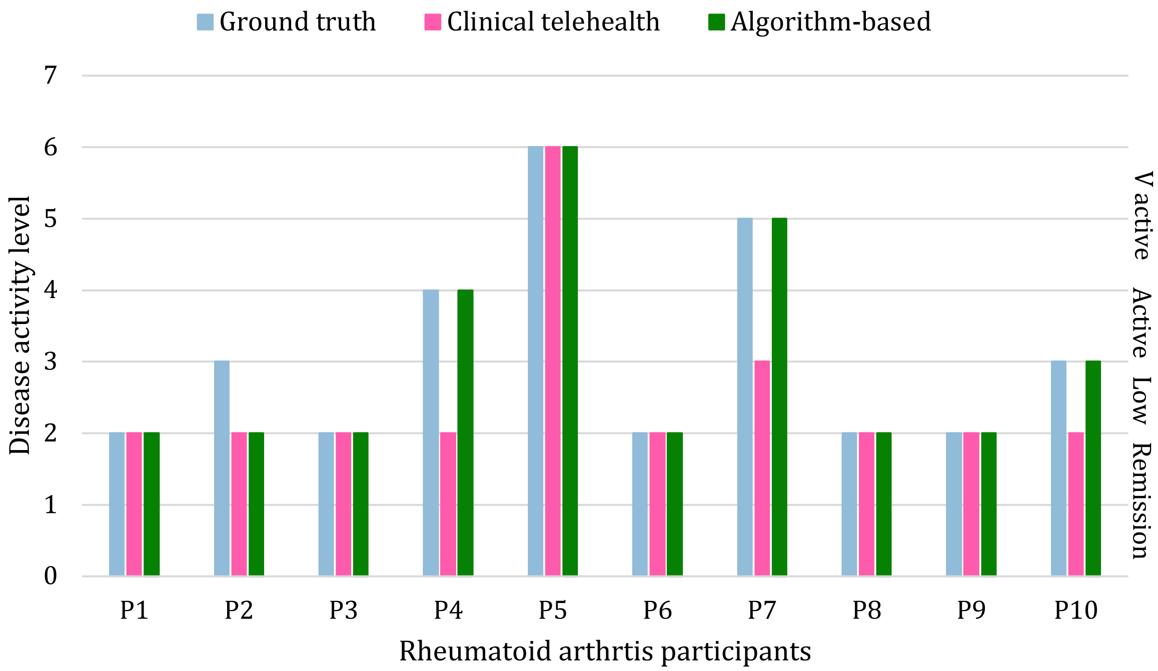
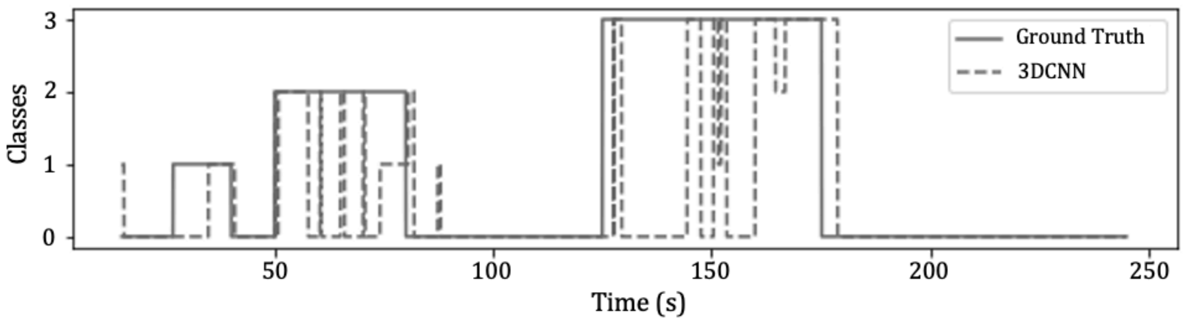
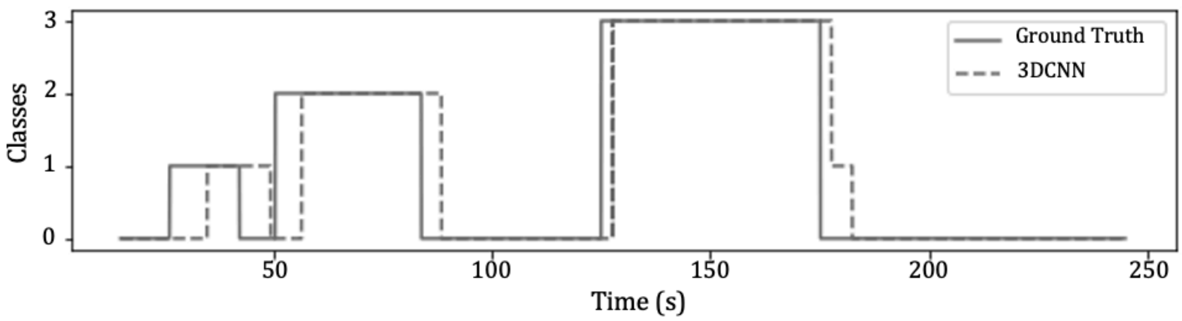


Figure 48: Bar chart representing categorical results obtained during the simulated clinical telehealth consultation (in pink) and the algorithm-based pipeline (in green) both compared against the ground truth (in blue) obtained in the clinic for the ten participants with rheumatoid arthritis (RA). The disease activity level goes shows RA patients in remission with low, active, and very active disease activity levels.

Results from the segmentation classifier demonstrated 84% (± 0.51) accuracy across the set with the mean Jaccard index $\overline{J_s}$ of 0.761 (± 0.2). The accuracy was uniformly distributed across all exercises, and all participants, reaching 89%, apart from participant 5 (with very active disease activity), where the segmentation results were 74%, the lowest predicted class scores. Illustrations of gesture segmentation for participant 5 (Figure 49A) and participant 1 in remission (Figure 49B) for the segmented activities show poor agreement for the participant with very active disease activity. In contrast, a better agreement between the manually segmented activities and the output from the three-dimensional CNN was presented for participant 1. Based on the segmentation output, for participant 5 a manual correction was implemented in around 7% of the entire dataset to ensure that OpenPose could be correctly executed on the segmented exercises.



(A)



(B)

Figure 49: Continuous temporal hand gesture segmentation and classification for the labels "no gesture" (class=0), "abduction and adduction" (class=1), "metacarpophalangeal (MCP) flexion" (class=2), and "thumb opposition"(class=3) for (A) patient number 5 with very active disease activity, and for (B) patient number 1 in remission. The ground truth (solid line) is compared against the three-dimensional convolutional neural network (3DCNN) results (dashed lines).

A representative plot for the MCP flexion activity (Figure 50) to illustrate the correspondence between MCP raw kinematics from healthy participants obtained in Chapter 4 and unhealthy kinematics gathered from RA patients. The plot illustrates how the average MCP values for twelve healthy participants obtained with OpenPose show higher ROMs than those for patients with low, active, and very active disease activity. From the values in Table 3 in Chapter 3, a ROM threshold was set at 50°, representing the maximum flexion value under which an MCP joint could be defined as having restricted movement.

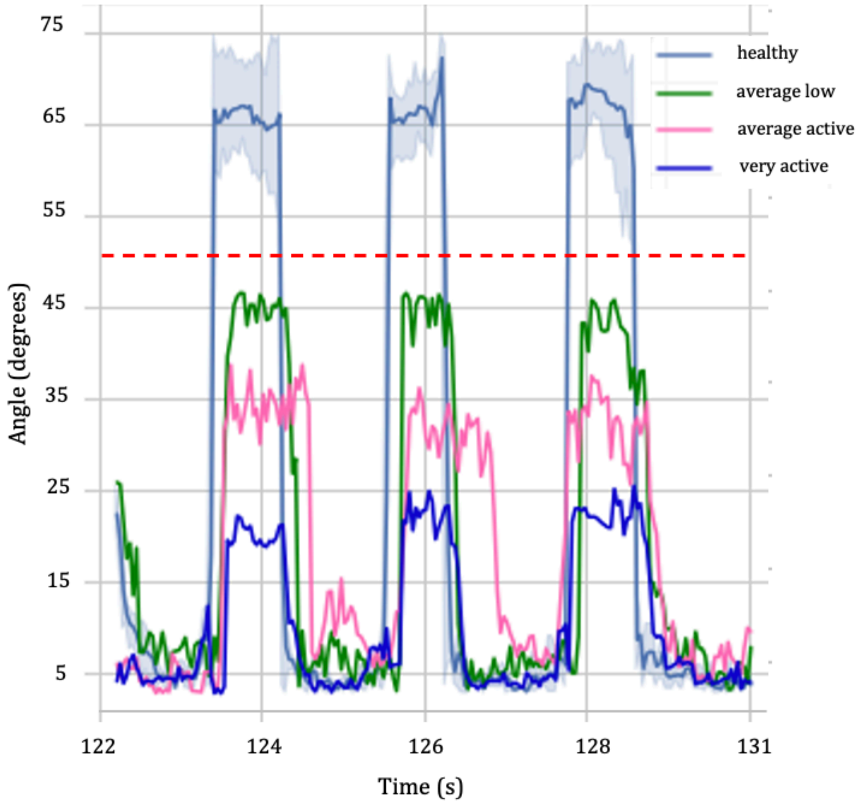
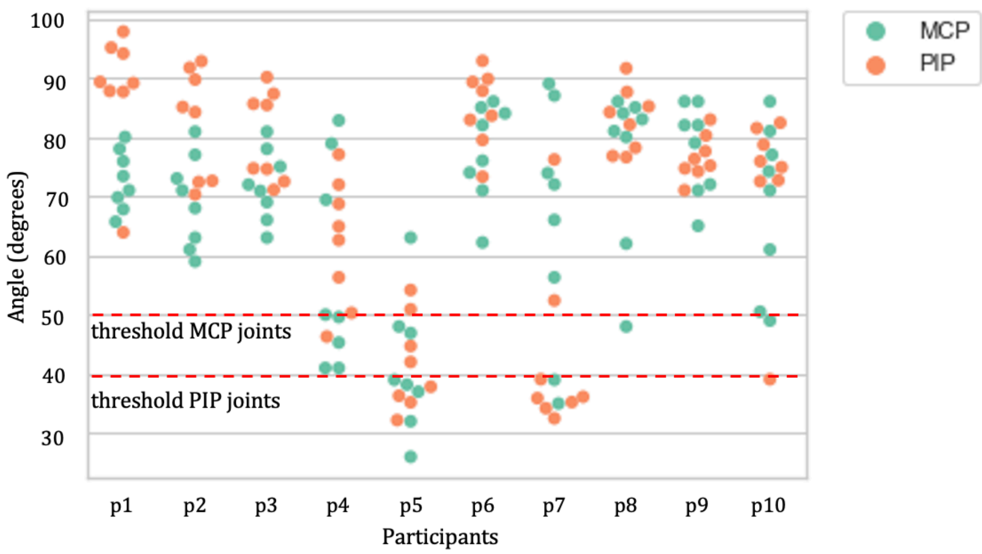
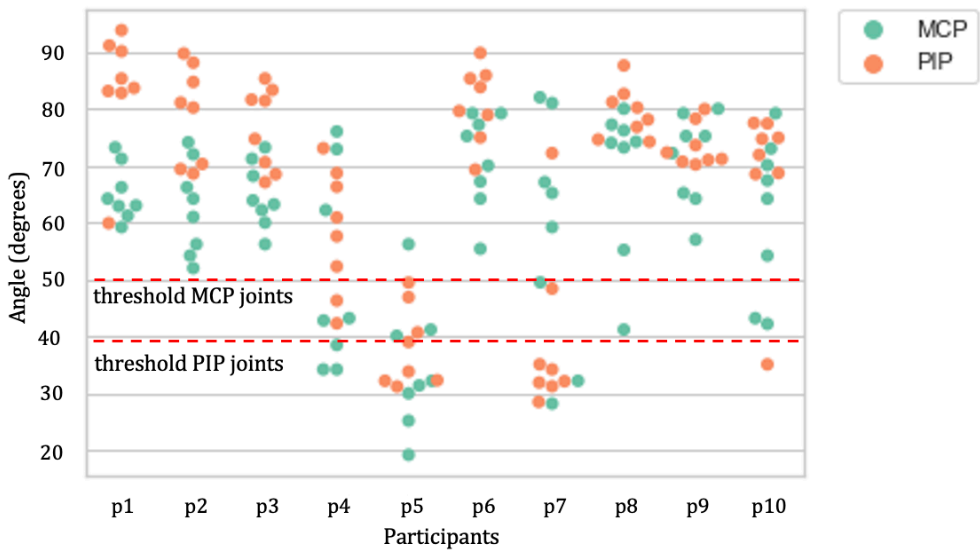


Figure 50: Raw data output from OpenPose during metacarpophalangeal flexion for twelve healthy participants (in blue; average +/- one standard deviation); average values of two participants with low disease activity (in green), average values of two participants with active disease activity (in pink); one participant with very active disease activity. Red dashed line indicates the maximum flexion value threshold under which the metacarpophalangeal joint is defined unhealthy.

Once the raw data were estimated, the TAF values were extracted to obtain the swollen joints and classify RA patients' disease activity levels (Figure 48). Bias values obtained in Chapter 4 for the left and right hands were applied to offset the TAF values and increase disease activity estimation. The biases were 4.72°, 6.82° and 4.70° for abduction-adduction, MCP flexion, and thumb opposition, respectively. Once applied, no difference was recorded in the disease activity status when offsetting the TAF values by the calculated bias, as illustrated for MCP and PIP joint angles for unbiased (Figure 51A) and biased (Figure 51B) data across participants. Results suggest that further investigations from a gold-standard marker-based motion capture system might be needed to find possible bias values specific to this population.



(A)



(B)

Figure 51: Scatterplots showing (A) unbiased and (B) biased total active flexion angles of the metacarpophalangeal (MCP) and proximal interphalangeal (PIP) joints during the metacarpophalangeal flexion and the thumb opposition activities.

The unbiased Cohen's kappa coefficient between the ground truth and the telehealth-based estimation was .4, indicating fair agreement between the two methods. The Cohen's kappa coefficient between the ground truth and algorithm-based estimations implemented in this chapter reported a value of .8, indicating substantial agreement between the two methods.

6.4 Discussion

Exploiting the pipeline implemented in Chapter 4 for the validation of OpenPose in tracking hand kinematics for specific hand exercises, in this third and final investigation, the validated keypoint detector was used to extract hand kinematics from video of RA participants performing hand exercises in front of a monocular RGB camera. To optimize the keypoint tracker and enable a smooth deployment in real-word applications, an additional segmentation classifier implemented in Chapter 5, was adopted with the aim of automatically detecting when hand activity occurred within a video sequence. Following a successful (84% of accuracy) segmentation of the classes, the keypoint detector automatically extracted the joint coordinates in each sub-set segmented, thereby enabling the extraction of MCP, PIP joint angles and kinematics of finger intersects. The values extracted were used to estimate disease activity. The inferred disease scores were compared with those obtained from video data visually marked by a rheumatologist and those assessed *apriori* during a face-to-face meeting in the clinic, which was designated as the ground truth.

The agreement between the ground truth and the algorithm-based estimation, illustrated in Figure 48, outperformed the rheumatologist’s assessment of the video. The categorical scores for the five RA participants in remission agreed with the scores provided by the two types of estimations, as participants exhibited a standard hand ROM. This result was anticipated as for RA patients in remission, the end goal reported in the therapeutic recommendations for the management of RA by the ACR/EULAR 2010 is for patients to maintain a full ROM (338). Normal ROM in the MCP, PIP joints, and angle intersects have also been reported in Mohammed et al.’s investigation of RA patients in remission (339).

Of the two RA patients with low disease activity (participants 2 and 10), the algorithm-based assessment was in agreement with the ground truth for participant 10, but not for participant 2, while the video consultation disagreed with the ground truth in both cases. The mocked video consultation also conflicted with the ground truth for RA patients diagnosed active disease activity (participants 4 and 7), while the algorithm-based approach picked up the limited ROM in both instances. These results are aligned with a cross-sectional investigation of the hand and wrist joints in rheumatoid patients suggesting that abnormality in the MCP and PIP joints is not easily identifiable by visual cues and necessitates objective quantification (260).

Finally, both (telehealth and algorithm-based) estimations were in agreement with the one participant (participant 5) with very active disease activity, showing extremely limited ROM. However, the sample size of two patients per category for active disease activity and one patient for very active disease activity are not representative and further research would be needed to enlarge the population and further validate the results.

Cohen's kappa was adopted as a measure of agreement between different forms of assessments; this is widely uses in estimating agreements in assessment in clinical evaluations that involve complex observations (260). This study found a 0.8 and a 0.4 agreement for the algorithm-based implementation and the simulated telehealth consultation (260). This illustrates that the proposed methodology based on convolutional units could outperform simulated telehealth assessments. Instead of using the Cohen's kappa index, some studies have suggested that the Fleiss' kappa could be considered to evaluate the agreement in studies where OpenPose is evaluated (340,341). However, it was not used in this study as evidence suggests that the Fleiss' kappa provides weak evidence on the significance agreement level.

The entire methodology presented in this chapter comes with some limitations that would need to be addressed in future research. The OpenPose workflow depends on a few post-processing steps, some of which were performed manually to clean up the dataset. For instance, manual post processing was done on the one subject where the segmentation provided 74% accuracy (Figure 49A). The proposed manual postprocessing steps suggest that further training for the three-dimensional CNN temporal classifier would be needed. Furthermore, unbiasing the data with the offset values observed in Chapter 4 did not correct the disagreement between the rheumatologist ratings and the algorithm-based ratings, emphasising the need for further data collection from an optoelectronic motion capture system, specific to an RA population.

It should be noted that simulated telehealth video consultations did not occur during an actual video call between RA patients and the rheumatologist assessor. The evaluation was only based on pre-recorded videos guided by the PowerPoint presentation, where the rheumatologist observed patients' hand motions. However, in a standard telemedicine consultation, the rheumatologist could have interactions with the patient, which would be missing in this case. However, this telehealth framework using pre-recorded asynchronous pre-recorded video recordings was suggested as the activity mainly during the COVID-19 pandemic to optimize rheumatology resources (336).

Furthermore, because the videos and the linked disease activity score utilized in this investigation were provided by a collaborator, relevant underlying features that would increase the understanding of the obtained results may have been lost. For instance, it was noted how many joints were swollen for each participant, but the ROMs for each joint obtained from clinical goniometric assessment were not provided. While objective assessment of the joint ROM can lead to improved access and more accurate outcomes, the current clinical practice involves providing a count of swollen joints based on visual examination. That said, the validated swollen joint count for RA has already been developed in several other clinical investigations. However, the kinematics gathered from video could have been compared with those obtained in clinic using goniometry-based assessment. Moreover, since the videos and ground truth disease activity were collected in a close time frame, but separately, in future investigations researchers could evaluate if data collected on the same visit would affect the accuracy of assessments.

Finally, it is anticipated that a larger dataset in a cross-sectional or longitudinal diagnostic clinical investigation containing identifiable information of RA participants can be used to provide more comprehensive results. Such an approach would enable a CNN model to be trained that would increase the validity of the study presented here. The robustness of such a methodology would include data collection from participants across different clinical centres, possibly considering retrospective data to predict disease activity trends.

This investigation provides an introductory exploration into the objective automated assessment of RA participants looking at hand kinematics. Results are encouraging, and provide preliminary evidence to assess how a vision-based technology can improve the quality of video consultations for the assessment of disease activity status for people with RA. This is desirable in the context of reducing the need for patients to travel and has been deemed to be particularly desirable given the recent experience of the COVID-19 pandemic. More broadly, this could improve patient outcomes through increased patient access to care, while offering optimization of healthcare resource utilization, including, but not limited to, reducing RA flares, urgent care and emergency room visits, and hospitalizations, all these aspects can be more fully explored in future studies.

Chapter 7 Discussion and conclusion

7.1 Introduction

Hand kinematics provide valuable metrics to assess and monitor healthy and impaired participants. Rheumatoid arthritis (RA), an autoimmune condition where the immune system affects the smaller joints first, particularly the fingers joints, can be estimated via monitoring the hands. RA demands a recurrent level of continuing observation, but face to face examinations are lengthy and remote consultation can lack objectivity.

In this thesis, attention has been given to deep-learning-based models that are able to infer from a single-coloured video recording the centre of the joint location. Such technology has been applied to impairment in gait, and this thesis aimed to 1) validate, 2) optimise, and 3) translate these models to hand function of RA patients.

This thesis has been developed through three key objectives:

1. to determine how accurately hand pose estimation models can evaluate human hand kinematic parameters, suggesting a pipeline for executing objective hand pose estimation examination using monocular RGB cameras.
2. to propose a temporal segmentation classifier methodology that differentiates between relevant hand gesture frames and transitional frames, combining both spatio-temporal motion parameters and appearance information.
3. to infer disease activity of patients with RA based on hand kinematics extracted using the validated markerless system, run on a subset of a continuous video sequence segmented using the temporal segmentation.

If proven successful, this leads to a validated pipeline that can support RA patients and clinicians to more frequently assess disease status and promote a low-cost management tool that can be used with RA patients to track and contain their disease through chronic and degenerative progression. Chapters 2 and 3 provide a review of systems for hand kinematics and remote monitoring tools used to capture clinical endpoints in RA. The evaluation of the accuracy assessment for the two-dimensional hand-oriented keypoint detector is then described in Chapter 4. Chapter 5 proposes a hand gesture recognition methodology from continuous video sequences. Finally, Chapter 6 validates the models in Chapter 4, and 5 on participants with hand issues due to RA.

7.2 Main findings

The next three sections summarise how this research addressed the objectives.

7.2.1 Objective 1: Two-dimensional video-based technology to track hand kinematics

Chapter 4 presented an open-source coloured video-based hand pose estimation model that delivers estimations of the centre of the joints of healthy human hands. The selected hand pose estimation model proposed in Chapter 4 needs only a two-dimensional digital video input to output spatiotemporal hand kinematic metrics. The key findings include:

i. The validated filtering approach previous proposed by Yao et al. (279) for lower limb kinematics extracted with OpenPose, extends to finger kinematics when using the same keypoint detector; the pipeline includes an Hampel filter with a window size of 4 for outlier removal followed by a Butterworth filter with a 3 HZ cut-off frequency.
ii. The markers introduced by the optoelectronic motion capture system (ground truth) did not influence the two-dimensional hand kinematics outputted by OpenPose.
iii. The root mean square error (RMSE) showed *acceptable* accuracy (defined based on goniometry measures to be below 10°) for the proposed hand activities, apart with a few outliers that occurred at a rate of 3.4% of the dataset.
iv. The coefficient of determination (R^2) of the linear regression between the total active flexion (TAF) values obtained from ground truth and OpenPose presented good agreement (above 0.7) for the abduction and adduction and the thumb

opposition tasks, and lower agreement ($0.4 < R^2 < 0.53$) for radial walking and the metacarpophalangeal (MCP) flexion activities.

v. The Bland-Altman plots illustrated that the mean differences between the ground truth and OpenPose ranged consistently and went from 4.7° to 6.8°.

Results indicated that OpenPose could be used as markerless technology to track hand kinematics in clinical studies, particularly when the MCP joints, the PIP joints, and the finger interests are under inspection.

7.2.2 Objective 2: Segmentation classifier

Chapter 5 offered an approach for large-scale video segmentation for hand gesture recognition. The video sequences were first segmented into single hand gesture sequences by classifying the frames into the different gestures. For one each of the segmented hand gesture series, the suggested technique utilized spatiotemporal information based on a three-dimensional convolutional neural network combined with a long-short-term memory unit. To enhance the accuracy of the model the training was performed on a large-scale hand dataset and fine-tuned for the relevant hand gestures with are part of the strategies to help RA participants do their hand exercises regularly.

Validation curves performed over for batch-sized 64 indicated good performances of the model, reaching an accuracy of 93.95% (± 0.37) with a mean Jaccard index of 0.812 (± 0.105) for a sample size of 22 participants. The presented model illustrated the possibility of training a model utilising a small set of data (113,410 fully labelled frames), compared to the more traditional convolutional neural networks that require vast labelling datasets. The presented pipeline adopted a small-sized architecture that could be executed to the acceptance of keypoint trackers and facilitate their adoption.

7.2.3 Objective 3: Proof of concept

Given the degenerative nature of RA, common illness signs include swelling, which causes decreased movement of the hands. Therefore, the assessment of swollen joints leads to a disease activity estimate, Chapter 6 shows the performances of adopting the approaches described in Chapters 4 and 5 to assess disease activity in individuals with RA.

The key findings include:

i. The clinically in-person assessed ground truth and the algorithm-based estimation agreed nine times out of ten. The asynchronous telehealth estimation was in accord with clinically in-person assessed ground truth six instances out of ten.
ii. The temporal segmentation model presented an accuracy of 84% (± 0.51 SD) with a mean Jaccard index of 0.761 (± 0.2) for the entire dataset. Furthermore, the accuracy was uniformly distributed across the different activities, reaching an accuracy of 89% for all participants, apart from the RA participant with very active disease activity, which presented an accuracy of 74%.
iii. The Cohen's kappa coefficient estimated between the in-clinic assessed ground truth and the asynchronous telehealth inspection was .4 (fair agreement). In contrast, the kappa coefficient between the in-person assessed ground truth and the algorithm-based estimations was .8 (substantial agreement).

Results support the evidence that objective automated assessment in RA can support disease activity fluctuations and enhance health care delivered during asynchronous video consultations.

7.3 Main strengths and limitations

This investigation has limitations, including the lack of tests under different visualization parameters and lightening conditions and the intrinsic inaccuracy of the tracking system (OpenPose). Also, the selected pre-trained network was chosen as previous studies had validated this model for lower limb kinematics. However, a pre-trained model was utilized, and this model was not trained for the specific hand exercises identified in the study.

Another limitation was identified by the extraction of two-dimensional hand keypoints, while the selected architecture (OpenPose) is also able to provide three-dimensional parameters when more than one camera is utilised. The difference in two-dimensional and three-dimensional parameters, as well as discrepancies in capturing the data from using different viewpoints or perspectives (e.g., sagittal, transverse), was not tested as part of this thesis due to the key objective of leveraging ubiquitous technology where users can upload videos using their own device.

In Chapter 6, the screening of the videos of RA participants performing hand exercises was performed verbally by the NRAS and this study might benefit to have more control over the captured clinical endpoints. For instance, playing with different recording techniques and different frame rates to evaluate the evaluate the possibility and the impact that the frame rate would have at capturing swollen joints. Also, the data collection was supplied at a single time-point. Such a clinical collection could have neglected the identification of a broader clinical representation (e.g., signs of illness progression over a long time).

A major limitation of the studies is the number of participants, which was particularly small for the clinical proof-of-concept. In addition, the classes of RA patients were unbalanced, with only one RA participant belonging to very high disease activity, for instance. For this reason, unbiasing the total active flexion values of the biases observed in Chapter 4 made no conclusive difference, and the results might have been more decisive if kinematics were captured from a gold-standard optoelectronic marker-based system using a larger and more balanced sample of patients. Particularly, the selection of participants was not randomized, and women were included more compared to men for the experiments presented in Chapter 4 and Chapter 6, which could limit the generalisation of the outcomes. Finally, it was not evaluated if the speed at which the hand exercises were performed could influence the tracking capabilities of the selected network.

7.4 Future research and overall conclusions

This thesis provides a pipeline that aim to lay the foundations for a remote management tool that can be used by participant suffering with RA. The system enables remote monitoring by using standard video cameras combined with two neural network models.

One model, an open-source model that tracks 21 keypoints of hand kinematics, is validated in Chapter 4 against an optoelectronic motion capture system; the second model, implemented in Chapter 5, enables the segmentation and classification of long gesture sequences. These models are then executed on video captures of users' hand/joint movements whilst performing hand exercises in Chapter 6. Here the networks segment and assess the ROM, estimating joint swelling and linking that to the disease progression with a disease activity score provided in output. The entire approach uses monocular cameras with the aim of leveraging ubiquitous technologies (e.g., in smartphones/laptops) and encourage the scalability of further investigations. Given the latest advantages of novel smartphone devices delivered with dual cameras, future investigations could include capturing image from additional cameras, enlarging the capabilities of this current investigation.

Future directions for the research include a longitudinal clinical investigation, e.g., one-year with monthly follow-ups, of this population that would furnish a further understanding into remote disease management. Looking at similar clinical investigations, one suggestion would be to enrol a larger, ethnically diverse population to broaden and generalise these findings. Future examinations could also include the incorporation of existing databases that look at the other components of the DAS-28 to implement other predictive algorithms that can work together with visual examination models to infer the disease activity. This could include electronic health records (EHR) data used as part of the current patient monitoring technology. The datasets include community-based retrospective datasets, such as the Norfolk Arthritis Register (342), the Scottish Early Rheumatoid Arthritis Study (343). With potential improvements made from assimilation of these datasets, this approach could not only be used to monitor patient's disease activity but would support potential prediction of future states of the patient, leading to forecast future clinical outcomes, as demonstrated by Norgeort et al. (344) in their preliminary investigation. This could enable better management of the disease, better-informed decisions on patients' treatment, and ultimately could improve RA patients' quality of life and lowering costs for medications.

With a system being capable of forecasting RA patient outcomes, or simulating potential outcomes under different treatment scenarios, the field of rheumatology could be positively impacted, with applications that go beyond the tracking of the condition, to include monitoring medication's effectiveness and helping to assess the need for secondary care. This would enable physicians to customise current therapeutic treatment plans to prevent RA from worsening and potentially avoid disease flare-ups requiring hospital care; ultimately, this would decrease hospitalisations, numbers of emergency department visits and the number of visits to general practitioners. Overall, with further investigations and larger clinical studies this would be a significant step towards improving the quality of care and optimising healthcare resources.

Bibliography

1. Chao EY. Biomechanics of the hand: a basic research study. World scientific; 1989.

2. Flanagan JR, Johansson RS. Hand movements. Encyclopedia of the human brain. 2002;2:399–414.

3. Martin LM. Clinical Mechanics of the Hand. American Journal of Occupational Therapy. 1988;42(3):199–199.

4. Carey MA, Laird DE, Murray KA, Stevenson JR. Reliability, validity, and clinical usability of a digital goniometer. Work. 2010;36(1):55–66.

5. Mayerson NH, Milano RA. Goniometric measurement reliability in physical medicine. Arch Phys Med Rehabil. 1984 Feb 1;65(2):92–4.

6. Bronner S, Agraharasamakulam S, Ojofeitimi S. Reliability and validity of electrogoniometry measurement of lower extremity movement. Journal of Medical Engineering & Technology. 2010 Apr;34(3):232–42.

7. Macionis V. Reliability of the standard goniometry and diagrammatic recording of finger joint angles: a comparative study with healthy subjects and non-professional raters. BMC Musculoskeletal Disorders. 2013 Jan 9;14(1):17.

8. Sturman DJ, Zeltzer D. A survey of glove-based input. IEEE Computer Graphics and Applications. 1994 Jan;14(1):30–9.

9. Dipietro L, Sabatini AM, Dario P. Evaluation of an instrumented glove for hand-movement acquisition. J Rehabil Res Dev. 2003 Apr;40(2):179–89.

10. Metcalf CD, Robinson R, Malpass AJ, Bogle TP, Dell TA, Harris C, et al. Markerless motion capture and measurement of hand kinematics: validation and application to home-based upper limb rehabilitation. IEEE Trans Biomed Eng. 2013 Aug;60(8):2184–92.

11. Mangal NK, Tiwari AK. A review of the evolution of scientific literature on technology-assisted approaches using RGB-D sensors for musculoskeletal health monitoring. Computers in Biology and Medicine. 2021;104316.

12. Cook JR, Baker NA, Cham R, Hale E, Redfern MS. Measurements of wrist and finger postures: a comparison of goniometric and motion capture techniques. Journal of applied biomechanics. 2007;23(1):70–8.

13. Metcalf CD. Challenges in measuring small joint movements: hand biomechanics and health technology assessment. 2009;

14. Guzsvinecz T, Szucs V, Sik-Lanyi C. Suitability of the Kinect Sensor and Leap Motion Controller—A Literature Review. Sensors (Basel). 2019 Mar 2;19(5):1072.

15. Yuan S, Ye Q, Stenger B, Jain S, Kim TK. Bighand2. 2m benchmark: Hand pose dataset and state of the art analysis. In 2017. p. 4866–74.

16. Zimmermann C, Brox T. Learning to estimate 3d hand pose from single rgb images. In 2017. p. 4903–11.

17. Cao Z, Hidalgo G, Simon T, Wei SE, Sheikh Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. IEEE Trans Pattern Anal Mach Intell. 2021 Jan 1;43(1):172–86.

18. Nakano N, Sakura T, Ueda K, Omura L, Kimura A, Iino Y, et al. Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. Frontiers in sports and active living. 2020;2:50.

19. Stenum J, Rossi C, Roemmich RT. Two-dimensional video-based analysis of human gait using pose estimation. PLoS Comput Biol. 2021 Apr 23;17(4):e1008935.

20. Sato K, Nagashima Y, Mano T, Iwata A, Toda T. Quantifying normal and parkinsonian gait features from home movies: Practical application of a deep learning–based 2D pose estimator. PloS one. 2019;14(11):e0223549.

21. Yao A, Gall J, Van Gool L. Coupled action recognition and pose estimation from multiple views. International journal of computer vision. 2012;100(1):16–37.

22. Zhu Y. Model-based human pose estimation with spatio-temporal inferencing. The Ohio State University; 2009.

23. Hoai M, Lan ZZ, De la Torre F. Joint segmentation and classification of human actions in video. In IEEE; 2011. p. 3265–72.

24. Wang P, Li W, Liu S, Zhang Y, Gao Z, Ogunbona P. Large-scale continuous gesture recognition using convolutional neural networks. In IEEE; 2016. p. 13–8.

25. Wang H. Two Stage Continuous Gesture Recognition Based on Deep Learning. Electronics. 2021;10(5):534.

26. Hakim NL, Shih TK, Kasthuri Arachchi SP, Aditya W, Chen YC, Lin CY. Dynamic hand gesture recognition using 3DCNN and LSTM with FSM context-aware model. Sensors. 2019;19(24):5429.

27. Zhu G, Zhang L, Shen P, Song J, Shah SAA, Bennamoun M. Continuous gesture segmentation and recognition using 3DCNN and convolutional LSTM. IEEE Transactions on Multimedia. 2018;21(4):1011–21.

28. Kidziński Ł, Yang B, Hicks JL, Rajagopal A, Delp SL, Schwartz MH. Deep neural networks enable quantitative movement analysis using single-camera videos. Nature communications. 2020;11(1):1–10.

29. Lonini L, Moon Y, Embry K, Cotton RJ, McKenzie K, Jenz S, et al. Video-Based Pose Estimation for Gait Analysis in Stroke Survivors during Clinical Assessments: A Proof-of-Concept Study. Digital Biomarkers. 2022;6(1):9–18.

30. Leo M, Bernava GM, Carcagnì P, Distante C. Video-Based Automatic Baby Motion Analysis for Early Neurological Disorder Diagnosis: State of the Art and Future Directions. Sensors. 2022;22(3):866.

31. Dixon WG, Michaud K. Using technology to support clinical care and research in rheumatoid arthritis. Current opinion in rheumatology. 2018;30(3):276.

32. Alamanos Y, Voulgari PV, Drosos AA. Rheumatoid arthritis in southern Europe: epidemiological, clinical, radiological and genetic considerations. Current Rheumatology Reviews. 2005;1(1):33–6.

33. Venables P, Maini RN. Diagnosis and differential diagnosis of rheumatoid arthritis. UpToDate2012. 2014;

34. de Jong Z, Munneke M, Zwinderman A, Kroon H, Ronday K, Lems W, et al. Long term high intensity exercise and damage of small joints in rheumatoid arthritis. Annals of the Rheumatic Diseases. 2004;63(11):1399–405.

35. Allen A, Carville S, McKenna F. Diagnosis and management of rheumatoid arthritis in adults: summary of updated NICE guidance. Bmj. 2018;362.

36. Smolen JS, Aletaha D. Monitoring rheumatoid arthritis. Current opinion in rheumatology. 2011;23(3):252–8.

37. Van Der Heijde D, Van Der Helm-Van AH, Aletaha D, Bingham CO, Burmester GR, Dougados M, et al. EULAR definition of erosive disease in light of the 2010 ACR/EULAR rheumatoid arthritis classification criteria. Annals of the rheumatic diseases. 2013;72(4):479–81.

38. Raad MW, Deriche MA, Hafeedh AB, Almasawa H, Jofan KB, Alsakkaf H, et al. An IOT based Wearable Smart Glove for Remote Monitoring of Rheumatoid Arthritis Patients. Biosignals. 2019;2019:224–8.

39. Henderson J, Condell J, Connolly J, Kelly D, Curran K. Review of Wearable Sensor-Based Health Monitoring Glove Devices for Rheumatoid Arthritis. Sensors. 2021;21(5):1576.

40. Cejnog LWX, de Campos T, Elui VMC, Cesar Jr. RM. A framework for automatic hand range of motion evaluation of rheumatoid arthritis patients. Informatics in Medicine Unlocked. 2021 Jan 1;23:100544.

41. Phutane U, Liphardt AM, Bräunig J, Penner J, Klebl M, Tascilar K, et al. Evaluation of Optical and Radar Based Motion Capturing Technologies for Characterizing Hand Movement in Rheumatoid Arthritis—A Pilot Study. Sensors. 2021;21(4):1208.

42. Austin L, Sharp CA, van der Veer SN, Machin M, Humphreys J, Mellor P, et al. Providing 'the bigger picture': benefits and feasibility of integrating remote monitoring from smartphones into the electronic health record: Findings from the Remote Monitoring of Rheumatoid Arthritis (REMORA) study. Rheumatology. 2020 Feb 1;59(2):367–78.

43. Hamy V, Garcia-Gancedo L, Pollard A, Myatt A, Liu J, Howland A, et al. Developing Smartphone-Based Objective Assessments of Physical Function in Rheumatoid Arthritis Patients: the PARADE Study. Digital biomarkers. 2020;4(1):26–44.

44. Miloslavsky EM, Bolster MB. Addressing the rheumatology workforce shortage: a multifaceted approach. In Elsevier; 2020. p. 791–6.

45. Roongta R, Ghosh A. Managing rheumatoid arthritis during COVID-19. Clinical rheumatology. 2020;1–8.

46. Bos WH, van Tubergen A, Vonkeman HE. Telemedicine for patients with rheumatic and musculoskeletal diseases during the COVID-19 pandemic; a positive experience in the Netherlands. Rheumatology international. 2021;41(3):565–73.

47. Donadoni F, Gionfrida L, Clarke B, Galloway J. P138 Developing a digital remote assessment for monitoring rheumatoid arthritis disease activity. Rheumatology. 2021;60(Supplement_1):keab247-134.

48. Van Riel P, Alten R, Combe B, Abdulganieva D, Bousquet P, Courtenay M, et al. Improving inflammatory arthritis management through tighter monitoring of patients and the use of innovative electronic tools. RMD open. 2016;2(2):e000302.

49. Kapandji IA. The physiology of the joints: annotated diagrams of the mechanics of the human joints. 2nd ed. London: E. & S. Livingstone; 1970. 3 p.

50. Pham T, Pathirana P, Trinh H, Fay P. A Non-Contact Measurement System for the Range of Motion of the Hand. Sensors. 2015 Jul 28;15(8):18315–33.

51. Hung LK, Ho KK, Leung PC. Impairment of hand function and loss of earning capacity after occupational hand injury: prospective cohort study. Hong Kong Med J. 1999 Sep;5(3):245–50.

52. Lutz W, Sanderson W, Scherbov S. The coming acceleration of global population ageing. Nature. 2008 Feb;451(7179):716–9.

53. Kandaswamy D, M M, Alexander M, Prabhu K, S MG, Krothapalli SB. Quantitative Assessment of Hand Dysfunction in Patients with Early Parkinson's Disease and Focal Hand Dystonia. J Mov Disord. 2018 Jan;11(1):35–44.

54. Yap HY, Tee S, Wong M, Chow SK, Peh SC, Teow SY. Pathogenic Role of Immune Cells in Rheumatoid Arthritis: Implications in Clinical Treatment and Biomarker Development. Cells. 2018 Oct 9;7(10):161.

55. Aristidou A, Lasenby J. Motion capture with constrained inverse kinematics for real-time hand tracking. In: 2010 4th International Symposium on Communications, Control and Signal Processing (ISCCSP) [Internet]. Limassol, Cyprus: IEEE; 2010 [cited 2021 Jul 15]. p. 1–5. Available from: http://ieeexplore.ieee.org/document/5463419/

56. Rath S. Hand kinematics: Application in clinical practice. Indian J Plast Surg. 2011;44(2):178–85.

57. ElKoura G, Singh K. Handrix: animating the human hand. In: Proceedings of the 2003 ACM SIGGRAPH/Eurographics symposium on Computer animation. Goslar, DEU: Eurographics Association; 2003. p. 110–9. (SCA '03).

58. Nanayakkara VK, Cotugno G, Vitzilaios N, Venetsanos D, Nanayakkara T, Sahinkaya MN. The role of morphology of the thumb in anthropomorphic grasping: a review. Frontiers in mechanical engineering. 2017;3:5.

59. Duncan SF, Saracevic CE, Kakinoki R. Biomechanics of the hand. Hand clinics. 2013;29(4):483–92.

60. Gates DH, Walters LS, Cowley J, Wilken JM, Resnik L. Range of Motion Requirements for Upper-Limb Activities of Daily Living. Am J Occup Ther. 2015 Dec 18;70(1):7001350010p1.

61. Gandbhir VN, Cunha B. Goniometer. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 [cited 2021 Jul 13]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK558985/

62. Rome K, Cowieson F. A Reliability Study of the Universal Goniometer, Fluid Goniometer, and Electrogoniometer for the Measurement of Ankle Dorsiflexion. Foot Ankle Int. 1996 Jan;17(1):28–32.

63. Sementille AC, Lourenço LE, Brega JRF, Rodello I. A motion capture system using passive markers. In: Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry - VRCAI '04 [Internet]. Singapore: ACM Press; 2004 [cited 2021 Aug 5]. p. 440. Available from: http://portal.acm.org/citation.cfm?doid=1044588.1044684

64. Yahya M, Shah JA, Kadir KA, Yusof ZM, Khan S, Warsi A. Motion capture sensing techniques used in human upper limb motion: a review. SR. 2019 Jul 15;39(4):504–11.

65. Field M, Pan Z, Stirling D, Naghdy F. Human motion capture sensors and analysis in robotics. Industrial Robot. 2011 Mar 8;38(2):163–71.

66. Kruk E van der, Reijne MM. Accuracy of human motion capture systems for sport applications; state-of-the-art review. European Journal of Sport Science. 2018 Jul 3;18(6):806–19.

67. Chatzis T, Stergioulas A, Konstantinidis D, Dimitropoulos K, Daras P. A Comprehensive Study on Deep Learning-Based 3D Hand Pose Estimation Methods. Applied Sciences. 2020 Sep 30;10(19):6850.

68. Sridhar S, Mueller F, Zollhöfer M, Casas D, Oulasvirta A, Theobalt C. Real-Time Joint Tracking of a Hand Manipulating an Object from RGB-D Input. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer Vision – ECCV 2016. Cham: Springer International Publishing; 2016. p. 294–310. (Lecture Notes in Computer Science).

69. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recognition. 2018;77:354–77.

70. Engstrand C, Krevers B, Kvist J. Interrater Reliability in Finger Joint Goniometer Measurement in Dupuytren's Disease. American Journal of Occupational Therapy. 2012 Jan 1;66(1):98–103.

71. Reese NB, Bandy WD, Yates C. Joint range of motion and muscle length testing. 2nd ed. St. Louis, Mo: Saunders/Elsevier; 2010. 509 p.

72. Colyer SL, Evans M, Cosker DP, Salo AIT. A Review of the Evolution of Vision-Based Motion Analysis and the Integration of Advanced Computer Vision Methods Towards Developing a Markerless System. Sports Med Open. 2018 Jun 5;4:24.

73. D'Antonio E, Taborri J, Palermo E, Rossi S, Patanè F. A markerless system for gait analysis based on OpenPose library. In: 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC). 2020. p. 1–6.

74. Drazan JF, Phillips WT, Seethapathi N, Hullfish TJ, Baxter JR. Moving outside the lab: markerless motion capture accurately quantifies sagittal plane kinematics during the vertical jump [Internet]. 2021 Mar [cited 2021 Sep 27] p. 2021.03.16.435503. Available from: https://www.biorxiv.org/content/10.1101/2021.03.16.435503v1

75. Noer HR, Pratt DR. A Goniometer Designed for the Hand. JBJS. 1958 Oct;40(5):1154–6.

76. Watklns MA, Riddle DL, Lamb RL, Personius WJ. Reliability of Goniometric Measurements and Visual Estimates of Knee Range of Motion Obtained in a Clinical Setting.

77. Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry. 4th ed. Philadelphia: F.A. Davis; 2009. 450 p.

78. Basic Concepts | Measurement of Joint Motion: A Guide to Goniometry, 5e | F.A. Davis PT Collection | McGraw Hill Medical [Internet]. [cited 2021 Jul 13]. Available from: https://fadavispt.mhmedical.com/Content.aspx?bookid=2124&sectionid=158980315

79. Gajdosik RL, Bohannon RW. Clinical measurement of range of motion. Review of goniometry emphasizing reliability and validity. Phys Ther. 1987 Dec;67(12):1867–72.

80. Vardakastani V, Bell H, Mee S, Brigstocke G, Kedgley AE. Clinical measurement of the dart throwing motion of the wrist: variability, accuracy and correction. J Hand Surg Eur Vol. 2018 Sep;43(7):723–31.

81. Kumar N, Pankaj D, Mahajan A, Kumar A, Sohi BS. Evaluation of normal Gait using electro-goniometer. JSIR Vol68(08) [August 2009] [Internet]. 2009 Aug [cited 2021 Jul 13]; Available from: http://nopr.niscair.res.in/handle/123456789/5302

82. Jonsson P, Johnson PW. Comparison of measurement accuracy between two types of wrist goniometer systems. Applied Ergonomics. 2001 Dec;32(6):599–607.

83. Ellis B, Bruton A. A study to compare the reliability of composite finger flexion with goniometry for measurement of range of motion in the hand. Clin Rehabil. 2002 Aug;16(5):562–70.

84. Marx RG, Bombardier C, Wright JG. What do we know about the reliability and validity of physical examination tests used to examine the upper extremity? The Journal of Hand Surgery. 1999 Jan;24(1):185–93.

85. Somers DL, Hanson JA, Kedzierski CM, Nestor KL, Quinlivan KY. The Influence of Experience on the Reliability of Goniometric and Visual Measurement of Forefoot Position. J Orthop Sports Phys Ther. 1997 Mar;25(3):192–202.

86. Laupattarakasem W, Sirichativapee W, Kowsuwon W, Sribunditkul S, Suibnugarn C. Axial Rotation Gravity Goniometer: A Simple Design of Instrument and a Controlled Reliability Study. Clinical Orthopaedics and Related Research. 1990 Feb;NA;(251):271???274.

87. Mermelstein MD, Blodgett JA. Single-mode optical fiber goniometer. Opt Lett. 1992 Jan 1;17(1):85.

88. Metcalf CD, Robinson R, Malpass AJ, Bogle TP, Dell TA, Harris C, et al. Markerless motion capture and measurement of hand kinematics: validation and application to home-based upper limb rehabilitation. IEEE Trans Biomed Eng. 2013 Aug;60(8):2184–92.

89. Yun Y, Agarwal P, Deshpande AD. Accurate, robust, and real-time estimation of finger pose with a motion capture system. In: 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2013. p. 1626–31.

90. Stancic I, Supuk TG, Panjkota A. Design, development and evaluation of optical motion-tracking system based on active white light markers. IET Science, Measurement & Technology. 2013;7(4):206–14.

91. van Andel CJ, Wolterbeek N, Doorenbosch CAM, Veeger DHEJ, Harlaar J. Complete 3D kinematics of upper extremity functional tasks. Gait Posture. 2008 Jan;27(1):120–7.

92. Williams KD, Ringer MA, Senior CA, Webb MJ, McAvaney BJ, Andronova N, et al. Evaluation of a component of the cloud response to climate change in an intercomparison of climate models. Clim Dyn. 2006 Feb;26(2–3):145–65.

93. [PDF] Inverse Kinematic Infrared Optical Finger Tracking | Semantic Scholar [Internet]. [cited 2021 Jul 14]. Available from: https://www.semanticscholar.org/paper/Inverse-Kinematic-Infrared-Optical-Finger-Tracking-Hillebrand-Bauer/4cacc809d92fd0c84d5b4b63d790dece998e0e6d

94. Maletsky LP, Sun J, Morton NA. Accuracy of an optical active-marker system to track the relative motion of rigid bodies. J Biomech. 2007;40(3):682–5.

95. Robertson: Research methods in biomechanics - Google Scholar [Internet]. [cited 2021 Jul 14]. Available from: https://scholar.google.com/scholar_lookup?title=Research%20Methods%20in%20Biomechanics&author=D.G.E.%20Robertson&publication_year=2014

96. Valevicius AM, Jun PY, Hebert JS, Vette AH. Use of optical motion capture for the analysis of normative upper body kinematics during functional upper limb tasks: A systematic review. J Electromyogr Kinesiol. 2018 Jun;40:1–15.

97. Yang N, Zhang M, Huang C, Jin D. Synergic analysis of upper limb target-reaching movements. J Biomech. 2002 Jun;35(6):739–46.

98. Anglin C, Wyss UP. Review of arm motion analyses. Proc Inst Mech Eng H. 2000;214(5):541–55.

99. Murphy MA, Sunnerhagen KS, Johnels B, Willén C. Three-dimensional kinematic motion analysis of a daily activity drinking from a glass: a pilot study. J NeuroEngineering Rehabil. 2006 Dec;3(1):18.

100. Cimolin V, Galli M, Albertini G, Crivellini M, Romkes J, Brunner R. Quantitative Analysis of Upper Limbs during Gait: A Marker Set Protocol. Journal of Applied Biomaterials & Functional Materials. 2012 Jan;10(1):49–55.

101. Della Croce U, Leardini A, Chiari L, Cappozzo A. Human movement analysis using stereophotogrammetry: Part 4: assessment of anatomical landmark misplacement and its effects on joint kinematics. Gait & Posture. 2005 Feb 1;21(2):226–37.

102. de Groot JH. The variability of shoulder motions recorded by means of palpation. Clinical Biomechanics. 1997 Oct 1;12(7):461–72.

103. Goislard de Monsabert B, Visser JMA, Vigouroux L, Van der Helm FCT, Veeger HEJ. Comparison of three local frame definitions for the kinematic analysis of the fingers and the wrist. Journal of Biomechanics. 2014 Aug 22;47(11):2590–7.

104. Manal K, McClay I, Stanhope S, Richards J, Galinat B. Comparison of surface mounted markers and attachment methods in estimating tibial rotations during walking: an in vivo study. Gait Posture. 2000 Feb;11(1):38–45.

105. Lu TW, Chang CF. Biomechanics of human movement and its clinical applications. The Kaohsiung Journal of Medical Sciences. 2012 Feb;28(2):S13–25.

106. Menache A. Understanding motion capture for computer animation. 2nd ed. Burlington, MA: Morgan Kaufmann; 2011. 254 p.

107. Poulin F, Amiot LP. Interference during the use of an electromagnetic tracking system under OR conditions. Journal of biomechanics. 2002;35(6):733–7.

108. Ehrig RM, Taylor WR, Duda GN, Heller MO. A survey of formal methods for determining the centre of rotation of ball joints. Journal of Biomechanics. 2006 Jan 1;39(15):2798–809.

109. Nakamura Y, Hanafusa H. Inverse Kinematic Solutions With Singularity Robustness for Robot Manipulator Control. Journal of Dynamic Systems, Measurement, and Control. 1986 Sep 1;108(3):163–71.

110. Wampler C. Manipulator Inverse Kinematic Solutions Based on Vector Formulations and Damped Least-Squares Methods. IEEE Trans Syst, Man, Cybern. 1986 Jan;16(1):93–101.

111. Aristidou A, Lasenby J. FABRIK: A fast, iterative solver for the Inverse Kinematics problem. Graphical Models. 2011;73(5):243–60.

112. Aristidou A, Lasenby J. Real-time marker prediction and CoR estimation in optical motion capture. The Visual Computer. 2013;29(1):7–26.

113. Rau G, Disselhorst-Klug C, Schmidt R. Movement biomechanics goes upwards: from the leg to the arm. Journal of Biomechanics. 2000 Oct 1;33(10):1207–16.

114. Chatzis T, Stergioulas A, Konstantinidis D, Dimitropoulos K, Daras P. A Comprehensive Study on Deep Learning-Based 3D Hand Pose Estimation Methods. undefined [Internet]. 2020 [cited 2021 Jul 14]; Available from: https://www.semanticscholar.org/paper/A-Comprehensive-Study-on-Deep-Learning-Based-3D-Chatzis-Stergioulas/b92831669677d259c6c2a1cc37ff14a3219a1486

115. Gabriella F, List R, Reissner L, Calcagni M. Development and Application of a Motion Analysis Protocol for the Kinematic Evaluation of Basic and Functional Hand and Finger Movements Using Motion Capture in a Clinical Setting—A Repeatability Study. Applied Sciences. 2020 Sep 16;10:6436.

116. Ferriero G, Sartorio F, Foti C, Primavera D, Brigatti E, Vercelli S. Reliability of a New Application for Smartphones (DrGoniometer) for Elbow Angle Measurement. PM&R. 2011 Dec;3(12):1153–4.

117. Jones A, Sealey R, Crowe M, Gordon S. Concurrent validity and reliability of the Simple Goniometer iPhone app compared with the Universal Goniometer. Physiotherapy Theory and Practice. 2014 Oct;30(7):512–6.

118. Ockendon M, Gilbert R. Validation of a Novel Smartphone Accelerometer-Based Knee Goniometer. J Knee Surg. 2012 May 3;25(04):341–6.

119. Otter SJ, Agalliu B, Baer N, Hales G, Harvey K, James K, et al. The reliability of a smartphone goniometer application compared with a traditional goniometer for measuring first metatarsophalangeal joint dorsiflexion. Journal of Foot and Ankle Research. 2015 Jul 23;8(1):30.

120. Alawna M, Unver B, Yuksel E. The Reliability of a Smartphone Goniometer Application Compared With a Traditional Goniometer for Measuring Ankle Joint

Range of Motion. Journal of the American Podiatric Medical Association. 2019 Jan 1;109:22–9.

121. The teardown: The Kinect for Xbox 360. Engineering & Technology. 2011 Apr 1;6(3):94–5.

122. Nizamis K, Rijken N, Mendes A, Janssen M, Bergsma A, Koopman B. A Novel Setup and Protocol to Measure the Range of Motion of the Wrist and the Hand. Sensors. 2018 Sep 25;18(10):3230.

123. Weichert F, Bachmann D, Rudak B, Fisseler D. Analysis of the accuracy and robustness of the leap motion controller. Sensors (Basel). 2013 May 14;13(5):6380–93.

124. Zhang Z. Microsoft kinect sensor and its effect. IEEE multimedia. 2012;19(2):4–10.

125. Shotton J, Fitzgibbon A, Cook M, Sharp T, Finocchio M, Moore R, et al. Real-Time Human Pose Recognition in Parts from Single Depth Images. :8.

126. Stone E, Skubic M. Evaluation of an Inexpensive Depth Camera for Passive In-Home Fall Risk Assessment. In 2011. p. 71–7.

127. Karbasi M, Bilal S, Aghababaeyan R, Rad AE, Bhatti Z, Shah A. ANALYSIS AND ENHANCEMENT OF THE DENOISING DEPTH DATA USING KINECT THROUGH ITERATIVE TECHNIQUE. In 2016.

128. Bachmann D, Weichert F, Rinkenauer G. Evaluation of the leap motion controller as a new contact-free pointing device. Sensors (Basel). 2014 Dec 24;15(1):214–33.

129. Niechwiej-Szwedo E, Gonzalez D, Nouredanesh M, Tung J. Evaluation of the Leap Motion Controller during the performance of visually-guided upper limb movements. PLOS ONE. 2018 Mar 12;13(3):e0193639.

130. Doosti B. Hand pose estimation: A survey. arXiv preprint arXiv:190301013. 2019;

131. Tompson J, Goroshin R, Jain A, LeCun Y, Bregler C. Efficient object localization using Convolutional Networks. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Boston, MA, USA: IEEE; 2015 [cited 2021 Jul 20]. p. 648–56. Available from: http://ieeexplore.ieee.org/document/7298664/

132. Toshev A, Szegedy C. DeepPose: Human Pose Estimation via Deep Neural Networks. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition [Internet]. Columbus, OH, USA: IEEE; 2014 [cited 2021 Jul 20]. p. 1653–60. Available from: https://ieeexplore.ieee.org/document/6909610

133. Wei SE, Ramakrishna V, Kanade T, Sheikh Y. Convolutional Pose Machines. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[Internet]. Las Vegas, NV, USA: IEEE; 2016 [cited 2021 Jul 20]. p. 4724–32. Available from: http://ieeexplore.ieee.org/document/7780880/

134. Sridhar S, Oulasvirta A, Theobalt C. Interactive markerless articulated hand motion tracking using RGB and depth data. In 2013. p. 2456–63.

135. Brownlee J. Deep learning for computer vision: image classification, object detection, and face recognition in python. Machine Learning Mastery; 2019.

136. Orozco CI, Buemi ME, Berlles JJ. CNN–LSTM Architecture for Action Recognition in Videos. :6.

137. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In 2015. p. 1–9.

138. Al-Hammadi M, Muhammad G, Abdul W, Alsulaiman M, Hossain MS. Hand Gesture Recognition Using 3D-CNN Model. IEEE Consumer Electronics Magazine. 2020 Jan;9(1):95–101.

139. Adithya V, Rajesh R. A deep convolutional neural network approach for static hand gesture recognition. Procedia Computer Science. 2020;171:2353–61.

140. Molchanov P, Gupta S, Kim K, Kautz J. Hand gesture recognition with 3D convolutional neural networks. In 2015. p. 1–7.

141. Flores CJL, Cutipa AG, Enciso RL. Application of convolutional neural networks for static hand gestures recognition under different invariant features. In IEEE; 2017. p. 1–4.

142. Ivanov YA, Bobick AF. Recognition of visual activities and interactions by stochastic parsing. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;22(8):852–72.

143. Psarrou A, Gong S, Walter M. Recognition of human gestures and behaviour based on motion trajectories. Image and Vision Computing. 2002;20(5–6):349–58.

144. De Mulder W, Bethard S, Moens MF. A survey on the application of recurrent neural networks to statistical language modeling. Computer Speech & Language. 2015;30(1):61–98.

145. A Guide to RNN: Understanding Recurrent Neural Networks and LSTM Networks [Internet]. Built In. [cited 2021 Aug 28]. Available from: https://builtin.com/data-science/recurrent-neural-networks-and-lstm

146. Shanthakumar VA, Peng C, Hansberger J, Cao L, Meacham S, Blakely V. Design and evaluation of a hand gesture recognition approach for real-time interactions. Multimed Tools Appl. 2020 Jul 1;79(25):17707–30.

147. Carreira J, Zisserman A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset. In: 2017 IEEE Conference on Computer Vision and Pattern

Recognition (CVPR) [Internet]. Honolulu, HI: IEEE; 2017 [cited 2021 Jul 19]. p. 4724–33. Available from: http://ieeexplore.ieee.org/document/8099985/

148. Mehta D, Sridhar S, Sotnychenko O, Rhodin H, Shafiei M, Seidel HP, et al. Vnect: Real-time 3d human pose estimation with a single rgb camera. ACM Transactions on Graphics (TOG). 2017;36(4):1–14.

149. Erol A, Bebis G, Nicolescu M, Boyle RD, Twombly X. Vision-based hand pose estimation: A review. Computer Vision and Image Understanding. 2007;108(1–2):52–73.

150. Chen L, Lin SY, Xie Y, Tang H, Xue Y, Xie X, et al. Generating realistic training images based on tonality-alignment generative adversarial networks for hand pose estimation. arXiv preprint arXiv:181109916. 2018;

151. Tome D, Russell C, Agapito L. Lifting from the Deep: Convolutional 3D Pose Estimation from a Single Image. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) [Internet]. Honolulu, HI: IEEE; 2017 [cited 2021 Jul 20]. p. 5689–98. Available from: http://ieeexplore.ieee.org/document/8100086/

152. Simon T, Joo H, Matthews I, Sheikh Y. Hand keypoint detection in single images using multiview bootstrapping. In 2017. p. 1145–53.

153. Miller NR, Shapiro R, McLaughlin TM. A technique for obtaining spatial kinematic parameters of segments of biomechanical systems from cinematographic data. Journal of Biomechanics. 1980;13(7):535–47.

154. Sakurai T, Okada H. EXAMINATION OF AN APPLICABLE RANGE FOR A MARKERLESS MOTION CAPTURE SYSTEM IN GAIT ANALYSIS. ISBS Proceedings Archive. 2021;39(1):141.

155. Sonkusare JS, Chopade NB, Sor R, Tade SL. A Review on Hand Gesture Recognition System. In: 2015 International Conference on Computing Communication Control and Automation. 2015. p. 790–4.

156. Aravind B, Ajay S, Priyadarshini. Virtualizing Electrical Appliances and Gadgets Input with Human Gestures. Procedia Computer Science. 2015 Jan 1;50:503–10.

157. Materzynska J, Berger G, Bax I, Memisevic R. The Jester Dataset: A Large-Scale Video Dataset of Human Gestures. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW) [Internet]. Seoul, Korea (South): IEEE; 2019 [cited 2021 Aug 10]. p. 2874–82. Available from: https://ieeexplore.ieee.org/document/9022297/

158. Allevard T, Benoit E, Foulloy L. Hand posture recognition with the fuzzy glove. In: Modern Information Processing. Elsevier; 2006. p. 417–27.

159. Primya T, Kanagaraj G, Muthulakshmi K, Chitra J, Gowthami A. Gesture recognition smart glove for speech impaired people. Materials Today:

Proceedings [Internet]. 2021 Feb 17 [cited 2021 Nov 25]; Available from: https://www.sciencedirect.com/science/article/pii/S2214785320405905

160. Ramamoorthy A, Vaswani N, Chaudhury S, Banerjee S. Recognition of dynamic hand gestures. Pattern Recognition. 2003;36(9):2069–81.

161. Pavlovic VI, Sharma R, Huang TS. Visual interpretation of hand gestures for human-computer interaction: A review. IEEE Transactions on pattern analysis and machine intelligence. 1997;19(7):677–95.

162. Wu W, Shi M, Wu T, Zhao D, Zhang S, Li J. Real-time Hand Gesture Recognition Based on Deep Learning in Complex Environments. In: 2019 Chinese Control And Decision Conference (CCDC) [Internet]. Nanchang, China: IEEE; 2019 [cited 2021 Jul 19]. p. 5950–5. Available from: https://ieeexplore.ieee.org/document/8833328/

163. Oudah M, Al-Naji A, Chahl J. Hand gesture recognition based on computer vision: a review of techniques. journal of Imaging. 2020;6(8):73.

164. Pisharady PK, Saerbeck M. Recent methods and databases in vision-based hand gesture recognition: A review. Computer Vision and Image Understanding. 2015;141:152–65.

165. Pinto RF, Borges CD, Almeida A, Paula IC. Static hand gesture recognition based on convolutional neural networks. Journal of Electrical and Computer Engineering. 2019;2019.

166. Gubrele P, Prasad R, Saurabh P, Verma B. Advance morphological filtering, correlation and convolution method for gesture recognition. In IEEE; 2017. p. 153–7.

167. Chen Q, Georganas ND, Petriu EM. Hand gesture recognition using Haar-like features and a stochastic context-free grammar. IEEE transactions on instrumentation and measurement. 2008;57(8):1562–71.

168. Castillo JC, Alonso-Martín F, Cáceres-Domínguez D, Malfaz M, Salichs MA. The Influence of Speed and Position in Dynamic Gesture Recognition for Human-Robot Interaction. Journal of Sensors. 2019;2019.

169. Liu S, Zhang J, Zhang Y, Zhu R. A wearable motion capture device able to detect dynamic motion of human limbs. Nat Commun. 2020 Nov 5;11(1):5615.

170. Bai R, Zhao Q, Zhou S, Li Y, Zhao X, Wang J. Continuous Action Recognition and Segmentation in Untrimmed Videos. In IEEE; 2018. p. 2534–9.

171. Porter S, Mirmehdi M, Thomas B. Temporal video segmentation and classification of edit effects. Image and Vision Computing. 2003;21(13–14):1097–106.

172. Darrell T, Pentland A. Space-time gestures. In IEEE; 1993. p. 335–40.

173. Wang JM, Fleet DJ, Hertzmann A. Gaussian process dynamical models. In Citeseer; 2005. p. 3.

174. Molchanov P, Yang X, Gupta S, Kim K, Tyree S, Kautz J. Online detection and classification of dynamic hand gestures with recurrent 3d convolutional neural network. In 2016. p. 4207–15.

175. Kuehne H, Gall J, Serre T. An end-to-end generative framework for video segmentation and recognition. In IEEE; 2016. p. 1–8.

176. Ni B, Yang X, Gao S. Progressively parsing interactional objects for fine grained action detection. In 2016. p. 1020–8.

177. Seethapathi N, Wang S, Saluja R, Blohm G, Kording KP. Movement science needs different pose tracking algorithms. arXiv preprint arXiv:190710226. 2019;

178. Nguyen TN, Huynh HH, Meunier J. Static hand gesture recognition using artificial neural network. Journal of Image and Graphics. 2013;1(1):34–8.

179. Islam MZ, Hossain MS, ul Islam R, Andersson K. Static hand gesture recognition using convolutional neural network with data augmentation. In IEEE; 2019. p. 324–9.

180. Galli M, Cimolin V, De Pandis MF, Schwartz MH, Albertini G. Use of the Gait Deviation index for the evaluation of patients with Parkinson's disease. Journal of motor behavior. 2012;44(3):161–7.

181. Seppen BF, L'ami MJ, dos Santos Rico SD, Ter Wee MM, Turkstra F, Roorda LD, et al. A smartphone app for self-monitoring of rheumatoid arthritis disease activity to assist patient-initiated care: protocol for a randomized controlled trial. JMIR research protocols. 2020;9(2):e15105.

182. Walker UA, Mueller RB, Jaeger VK, Theiler R, Forster A, Dufner P, et al. Disease activity dynamics in rheumatoid arthritis: patients' self-assessment of disease activity via WebApp. Rheumatology. 2017;56(10):1707–12.

183. White KM, Ivan A, Williams R, Galloway JB, Norton S, Matcham F. Remote Measurement in Rheumatoid Arthritis: Qualitative Analysis of Patient Perspectives. JMIR formative research. 2021;5(3):e22473.

184. Rodríguez-Sánchez-Laulhé P, Luque-Romero LG, Blanquero J, Suero-Pineda A, Biscarri-Carbonero Á, Barrero-García FJ, et al. A mobile app using therapeutic exercise and education for self-management in patients with hand rheumatoid arthritis: a randomized controlled trial protocol. Trials. 2020;21(1):1–12.

185. Salaffi F, Carotti M, Ciapetti A, Di Carlo M, Gasparini S, Farah S, et al. Effectiveness of a telemonitoring intensive strategy in early rheumatoid arthritis: comparison with the conventional management approach. BMC musculoskeletal disorders. 2016;17(1):1–11.

186. England B, Mikuls TR. Epidemiology of, risk factors for, and possible causes of rheumatoid arthritis. 2020;

187. Grassi W, De Angelis R, Lamanna G, Cervini C. The clinical features of rheumatoid arthritis. European Journal of Radiology. 1998 May;27:S18–24.

188. Williams MA, Williamson EM, Heine PJ, Nichols V, Glover MJ, Dritsaki M, et al. Strengthening And stretching for Rheumatoid Arthritis of the Hand (SARAH). A randomised controlled trial and economic evaluation. Health Technology Assessment. 2015 Mar;19(19):1–222.

189. Van der Heijde D, van't Hof MA, Van Riel P, Theunisse L, Lubberts EW, van Leeuwen MA, et al. Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. Annals of the rheumatic diseases. 1990;49(11):916–20.

190. Smolen JS, Landewé RB, Bijlsma JW, Burmester GR, Dougados M, Kerschbaumer A, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2019 update. Annals of the rheumatic diseases. 2020;79(6):685–99.

191. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for People With Rheumatoid Arthritis to Monitor Their Disease Activity: A Review of Apps for Best Practice and Quality. JMIR Mhealth Uhealth. 2017 Feb 21;5(2):e7.

192. Overview | Rheumatoid arthritis in adults: management | Guidance | NICE [Internet]. NICE; [cited 2021 Aug 16]. Available from: https://www.nice.org.uk/guidance/ng100

193. Hammond A, Prior Y. The effectiveness of home hand exercise programmes in rheumatoid arthritis: a systematic review. British medical bulletin. 2016;119(1):49–62.

194. Cooney JK, Law RJ, Matschke V, Lemmey AB, Moore JP, Ahmad Y, et al. Benefits of exercise in rheumatoid arthritis. Journal of aging research. 2011;2011.

195. Williams MA, Srikesavan C, Heine PJ, Bruce J, Brosseau L, Hoxey-Thomas N, et al. Exercise for rheumatoid arthritis of the hand. Cochrane Database of Systematic Reviews. 2018;(7).

196. He M, Zhou W, Guo J, Liu J, Dong C, Zhao R, et al. AB0146 HAND AND WRIST ACTIVE RANGE OF MOTIONS AND CONTRIBUTING FACTORS IN RHEUMATOID ARTHRITIS: A CROSS-SECTIONAL STUDY. 2021;

197. Pani D, Barabino G, Dessi A, Tradori I, Piga M, Mathieu A, et al. A device for local or remote monitoring of hand rehabilitation sessions for rheumatic patients. IEEE journal of translational engineering in health and medicine. 2014;2:1–11.

198. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for

managing health: systematic review and content analysis of findings. Journal of medical Internet research. 2018;20(7):e10480.

199. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. Arthritis & Rheumatism. 2008;58(1):15–25.

200. Myasoedova E, Gabriel SE, Matteson EL, Davis JM, Therneau TM, Crowson CS. Decreased cardiovascular mortality in patients with incident rheumatoid arthritis (RA) in recent years: dawn of a new era in cardiovascular disease in RA? The Journal of rheumatology. 2017;44(6):732–9.

201. Dadoun S, Zeboulon-Ktorza N, Combescure C, Elhai M, Rozenberg S, Gossec L, et al. Mortality in rheumatoid arthritis over the last fifty years: systematic review and meta-analysis. Joint Bone Spine. 2013;80(1):29–33.

202. Symmons D, Turner G, Webb R, Asten P, Barrett E, Lunt M, et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology. 2002;41(7):793–800.

203. Burton W, Morrison A, Maclean R, Ruderman E. Systematic review of studies of productivity loss due to rheumatoid arthritis. Occupational Medicine. 2006;56(1):18–27.

204. Agarwal SK. Core management principles in rheumatoid arthritis to help guide managed care professionals. Journal of Managed Care Pharmacy. 2011;17(9 Supp B):S03–8.

205. Bowman SJ, Guest L. The national clinical audit for rheumatoid and early inflammatory arthritis. Clin Med (Lond). 2016 Dec;16(6):500–1.

206. Birnbaum H, Pike C, Kaufman R, Marynchenko M, Kidolezi Y, Cifaldi M. Societal cost of rheumatoid arthritis patients in the US. Curr Med Res Opin. 2010 Jan;26(1):77–90.

207. Scott D. Prognostic factors in early rheumatoid arthritis. Rheumatology (Oxford, England). 2000;39:24–9.

208. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. New England Journal of Medicine. 2011;365(23):2205–19.

209. Dantzer R, Capuron L, Irwin MR, Miller AH, Ollat H, Perry VH, et al. Identification and treatment of symptoms associated with inflammation in medically ill patients. Psychoneuroendocrinology. 2008 Jan;33(1):18–29.

210. Donate PB, de Lima KA, Peres RS, Almeida F, Fukada SY, Silva TA, et al. Cigarette smoke induces miR-132 in Th17 cells that enhance osteoclastogenesis in inflammatory arthritis. Proceedings of the National Academy of Sciences. 2021;118(1).

211. Goia DN, Fortulan CA, Purquerio BM, Elui VMC. A new concept of orthosis for correcting fingers ulnar deviation. Research on Biomedical Engineering. 2017;33:050–7.

212. Weissman BN, Aliabadi P, Weinfeld MS, Thomas WH, Sosman JL. Prognostic features of atlantoaxial subluxation in rheumatoid arthritis patients. Radiology. 1982 Sep;144(4):745–51.

213. Chung KC, Pushman AG. Current Concepts in the Management of the Rheumatoid Hand. The Journal of Hand Surgery. 2011 Apr;36(4):736–47.

214. Sebastin SJ, Chung KC. Reconstruction of digital deformities in rheumatoid arthritis. Hand clinics. 2011;27(1):87.

215. Fornaciari G, Fontecchio G, Ventura L, Papola F, Trombetta I, Giuffra V. Rheumatoid arthritis in Cardinal Carlo de'Medici (1595-1666): A confirmed macroscopic, radiologic and molecular diagnosis. Clinical and Experimental Rheumatology-Incl Supplements. 2012;30(1):12.

216. Towheed TE, Anastassiades TP. Rheumatoid hand. Practical approach to assessment and management. Can Fam Physician. 1994 Jul;40:1303–9.

217. Ansell B. Hand assessment in rheumatoid arthritis. Annals of the rheumatic diseases. 1969;28(Suppl 5):77.

218. Levy G, Cheetham C, Cheatwood A, Burchette R. Validation of patient-reported joint counts in rheumatoid arthritis and the role of training. The Journal of rheumatology. 2007;34(6):1261–5.

219. Van der Linden M, Knevel R, Huizinga T, van der Helm-van Mil AM. Classification of rheumatoid arthritis: comparison of the 1987 American College of Rheumatology criteria and the 2010 American College of Rheumatology/European League Against Rheumatism criteria. Arthritis & Rheumatism. 2011;63(1):37–42.

220. Carpenter L. Long-term Progression of Structural Joint Damage in Early Rheumatoid Arthritis. 2017;

221. Mota LMH da, Laurindo IMM, Santos Neto LL dos, Lima FAC, Viana SL, Mendlovitz PS, et al. Imaging diagnosis of early rheumatoid arthritis. Revista brasileira de reumatologia. 2012;52:761–6.

222. Kgoebane K, Ally MM, Duim-Beytell MC, Suleman FE. The role of imaging in rheumatoid arthritis. SA journal of radiology. 2018;22(1):1–6.

223. Prevoo M, Van'T Hof M, Kuper H, Van Leeuwen M, Van De Putte L, Van Riel P. Modified disease activity scores that include twenty-eight-joint counts development and validation in a prospective longitudinal study of patients with rheumatoid arthritis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 1995;38(1):44–8.

224. Scott D, Houssien D. Joint assessment in rheumatoid arthritis. Rheumatology. 1996;35(suppl_2):14–8.

225. Firestein GS, Budd RC, Gabriel SE, McInnes IB, O'Dell JR. Kelley and Firestein's textbook of rheumatology e-book. Elsevier Health Sciences; 2016.

226. Essers I, Boonen A, Busch M, van der Heijde D, Keszei AP, Landewé R, et al. Fluctuations in patient reported disease activity, pain and global being in patients with ankylosing spondylitis. Rheumatology. 2016 Nov 1;55(11):2014–22.

227. Smolen JS, Szumski A, Koenig AS, Jones TV, Marshall L. Predictors of remission with etanercept-methotrexate induction therapy and loss of remission with etanercept maintenance, reduction, or withdrawal in moderately active rheumatoid arthritis: results of the PRESERVE trial. Arthritis research & therapy. 2018;20(1):1–17.

228. Wells G, Becker J, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. Annals of the rheumatic diseases. 2009;68(6):954–60.

229. Koevoets R, Allaart CF, van der HEIJDE DM, Huizinga TW. Disease activity monitoring in rheumatoid arthritis in daily practice: experiences with METEOR, a free online tool. The Journal of Rheumatology. 2010;37(12):2632–3.

230. Ritchie DM, Boyle JA, MCINNES JM, Jasani MK, Dalakos TG, Grieveson P, et al. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. QJM: An International Journal of Medicine. 1968;37(3):393–406.

231. Van Gestel A, Prevoo M, Van't Hof M, Van Rijswijk M, Van de Putte L, Van Riel P. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis: comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism criteria. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology. 1996;39(1):34–40.

232. Fransen J, Creemers M, Van Riel P. Remission in rheumatoid arthritis: agreement of the disease activity score (DAS28) with the ARA preliminary remission criteria. Rheumatology. 2004;43(10):1252–5.

233. Solomon DH, Bitton A, Katz JN, Radner H, Brown E, Fraenkel L. Treat to target in rheumatoid arthritis: fact, fiction or hypothesis? Arthritis & rheumatology (Hoboken, NJ). 2014;66(4):775.

234. Goekoop-Ruiterman Y d, de Vries-Bouwstra J, Allaart C, Van Zeben D, Kerstens P, Hazes J, et al. Clinical and radiographic outcomes of four different treatment

strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. Arthritis & Rheumatism. 2005;52(11):3381–90.

235. Landewé RB, Boers M, Verhoeven AC, Westhovens R, van de Laar MA, Markusse HM, et al. COBRA combination therapy in patients with early rheumatoid arthritis: long-term structural benefits of a brief intervention. Arthritis & Rheumatism. 2002;46(2):347–56.

236. Services for people with rheumatoid arthritis - National Audit Office (NAO) Press release [Internet]. National Audit Office. 2009 [cited 2021 Aug 16]. Available from: https://www.nao.org.uk/press-release/services-for-people-with-rheumatoid-arthritis-2/

237. Cush J. Biological drug use: US perspectives on indications and monitoring. Annals of the rheumatic diseases. 2005;64(suppl 4):iv18–23.

238. Wolfe F, Pincus T, Thompson AK, Doyle J. The assessment of rheumatoid arthritis and the acceptability of self-report questionnaires in clinical practice. Arthritis Care & Research: Official Journal of the American College of Rheumatology. 2003;49(1):59–63.

239. Guo Q, Wang Y, Xu D, Nossent J, Pavlos NJ, Xu J. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. Bone research. 2018;6(1):1–14.

240. Wasserman A. Diagnosis and management of rheumatoid arthritis. American family physician. 2011;84(11):1245–52.

241. Figueroa F, Braun-Moscovici Y, Khanna D, Voon E, Gallardo L, Luinstra D, et al. Patient self-administered joint tenderness counts in rheumatoid arthritis are reliable and responsive to changes in disease activity. The Journal of rheumatology. 2007;34(1):54–6.

242. Singh H, Kumar S, Talapatra P, Gupta V, Ray S, Kumar H. Assessment of hand functions in rheumatoid arthritis using SF-SACRAH (short form score for the assessment and quantification of chronic rheumatoid affections of the hands) and its correlation to disease activity. Rheumatology international. 2012;32(11):3413–9.

243. Bergstra S, Murgia A, Te Velde A, Caljouw S. A systematic review into the effectiveness of hand exercise therapy in the treatment of rheumatoid arthritis. Clinical rheumatology. 2014;33(11):1539–48.

244. Renskers L, van Uden RJ, Huis AM, Rongen SA, Teerenstra S, van Riel PL. Comparison of the construct validity and reproducibility of four different types of patient-reported outcome measures (PROMs) in patients with rheumatoid arthritis. Clinical rheumatology. 2018;37(12):3191–9.

245. Chaudhari P. The impact of rheumatoid arthritis and biologics on employers and payers. Biotechnology healthcare. 2008;5(2):37.

246. Lamb SE, Williamson EM, Heine PJ, Adams J, Dosanjh S, Dritsaki M, et al. Exercises to improve function of the rheumatoid hand (SARAH): a randomised controlled trial. The Lancet. 2015;385(9966):421–9.

247. Hoenig H, Groff G, Pratt K, Goldberg E, Franck W. A randomized controlled trial of home exercise on the rheumatoid hand. The Journal of rheumatology. 1993;20(5):785–9.

248. Sherrington C, Herbert R, Maher C, Moseley A. PEDro. A database of randomized trials and systematic reviews in physiotherapy. Manual therapy. 2000;5(4):223–6.

249. Manning VL, Hurley MV, Scott DL, Coker B, Choy E, Bearne LM. Education, self-management, and upper extremity exercise training in people with rheumatoid arthritis: a randomized controlled trial. Arthritis care & research. 2014;66(2):217–27.

250. Roddy E, Zhang W, Doherty M, Arden NK, Barlow J, Birrell F, et al. Evidence-based recommendations for the role of exercise in the management of osteoarthritis of the hip or knee--the MOVE consensus. Rheumatology (Oxford). 2005 Jan;44(1):67–73.

251. Heine P, Williams M, Williamson E, Bridle C, Adams J, O'Brien A, et al. Development and delivery of an exercise intervention for rheumatoid arthritis: strengthening and stretching for rheumatoid arthritis of the hand (SARAH) trial. Physiotherapy. 2012;98(2):121–30.

252. O'Brien AV, Jones P, Mullis R, Mulherin D, Dziedzic K. Conservative hand therapy treatments in rheumatoid arthritis—a randomized controlled trial. Rheumatology. 2006 May 1;45(5):577–83.

253. Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, et al. Using a ResearchKit smartphone app to collect rheumatoid arthritis symptoms from real-world participants: feasibility study. JMIR mHealth and uHealth. 2018;6(9):e177.

254. Reade S, Spencer K, Sergeant JC, Sperrin M, Schultz DM, Ainsworth J, et al. Cloudy with a chance of pain: engagement and subsequent attrition of daily data entry in a smartphone pilot study tracking weather, disease severity, and physical activity in patients with rheumatoid arthritis. JMIR mHealth and uHealth. 2017;5(3):e37.

255. Pons JL, Torricelli D. Emerging therapies in neurorehabilitation. Springer; 2014.

256. Marengo MF, Suarez-Almazor ME. Improving treatment adherence in patients with rheumatoid arthritis: what are the options? International journal of clinical rheumatology. 2015;10(5):345.

257. Smolen JS, Landewé R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Annals of the rheumatic diseases. 2010;69(6):964–75.

258. Majithia V, Geraci SA. Rheumatoid arthritis: diagnosis and management. The American journal of medicine. 2007;120(11):936–9.

259. Goossens P, Heemskerk B, Van Tongeren J, Zwinderman A, Vlieland TV, Huizinga T. Reliability and sensitivity to change of various measures of hand function in relation to treatment of synovitis of the metacarpophalangeal joint in rheumatoid arthritis. Rheumatology. 2000;39(8):909–13.

260. Zhang L, Cao H, Zhang Q, Fu T, Yin R, Xia Y, et al. Motion analysis of the wrist joints in Chinese rheumatoid arthritis patients: a cross-sectional study. BMC musculoskeletal disorders. 2018;19(1):1–8.

261. Turner DE, Woodburn J. Characterising the clinical and biomechanical features of severely deformed feet in rheumatoid arthritis. Gait & posture. 2008;28(4):574–80.

262. Aleixo P, Vaz Patto J, Cardoso A, Moreira H, Abrantes J. Ankle kinematics and kinetics during gait in healthy and rheumatoid arthritis post-menopausal women. Somatosensory & motor research. 2019;36(2):171–8.

263. Soucie J, Wang C, Forsyth A, Funk S, Denny M, Roach K, et al. Range of motion measurements: reference values and a database for comparison studies. Haemophilia. 2011;17(3):500–7.

264. Adams J, Cavolina J, Read H, Larkin E. Differences in hand functional ability in men and women with and without rheumatoid arthritis. In BMJ publishing group BMA; 2007. p. 658–658.

265. Horsten NC, Ursum J, Roorda LD, van Schaardenburg D, Dekker J, Hoeksma AF. Prevalence of hand symptoms, impairments and activity limitations in rheumatoid arthritis in relation to disease duration. Journal of Rehabilitation Medicine. 2010;42(10):916–21.

266. Epps H, Hurley M, Utley M. Development and evaluation of a single value score to assess global range of motion in juvenile idiopathic arthritis. Arthritis Care & Research. 2002;47(4):398–402.

267. Gutiérrez-Martínez J, Ortiz-Espinosa A, Hernández-Rodríguez PR, Núñez-Gaona MA. System to measure the range of motion of the joints of the human hand. Revista de Investigacion Clinica; Organo del Hospital de Enfermedades de la Nutricion. 2014;66:S122-30.

268. Hasselkus BR, Kshepakaran K, Safrit MJ. Handedness and hand joint changes in rheumatoid arthritis. The American Journal of Occupational Therapy. 1981;35(11):705–10.

269. Highton J, Solomon C, Gardiner D, Doyle T. Video image analysis of hands: development of an 'anatomic index'as a potential outcome measure in rheumatoid arthritis. Rheumatology. 1996;35(12):1274–80.

270. Meals CG, Saunders RJ, Desale S, Means Jr KR. Viability of hand and wrist photogoniometry. Hand. 2018;13(3):301–4.

271. Lima L, Melo J, Fragoso T, Vieira T, Oliveira M. Fisiomotion: Sistema de avaliação de pacientes portadores de artrite reumatoide usando sensor de movimentos. In 2016.

272. Ganguly A, Rashidi G, Mombaur K. Comparison of the Performance of the Leap Motion ControllerTM with a Standard Marker-Based Motion Capture System. Sensors. 2021;21(5):1750.

273. Chen X, Wang G, Guo H, Zhang C. Pose guided structured region ensemble network for cascaded hand pose estimation. Neurocomputing. 2020;395:138–49.

274. Metcalf CD, Notley SV, Chappell PH, Burridge JH, Yule VT. Validation and Application of a Computational Model for Wrist and Hand Movements Using Surface Markers. IEEE Trans Biomed Eng. 2008 Mar;55(3):1199–210.

275. Lopes TJA, Ferrari D, Ioannidis J, Simic M, Mícolis de Azevedo F, Pappas E. Reliability and validity of frontal plane kinematics of the trunk and lower extremity measured with 2-dimensional cameras during athletic tasks: A systematic review with meta-analysis. journal of orthopaedic & sports physical therapy. 2018;48(10):812–22.

276. Reinking MF, Dugan L, Ripple N, Schleper K, Scholz H, Spadino J, et al. Reliability of two-dimensional video-based running gait analysis. International journal of sports physical therapy. 2018;13(3):453.

277. Hensley CP, Millican D, Hamilton N, Yang A, Lee J, Chang AH. Video-based motion analysis use: A national survey of orthopedic physical therapists. Physical therapy. 2020;100(10):1759–70.

278. Owoeye OB, Rauvola RS, Brownson RC. Dissemination and implementation research in sports and exercise medicine and sports physical therapy: translating evidence to practice and policy. BMJ Open Sport & Exercise Medicine. 2020;6(1):e000974.

279. Buljina AI, Taljanovic MS, Avdic DM, Hunter TB. Physical and exercise therapy for treatment of the rheumatoid hand. Arthritis Care & Research: Official Journal of the American College of Rheumatology. 2001;45(4):392–7.

280. Metcalf CD, Notley SV. Modified kinematic technique for measuring pathological hyperextension and hypermobility of the interphalangeal joints. IEEE Trans Biomed Eng. 2011 May;58(5):1224–31.

281. Smith S. Digital signal processing: a practical guide for engineers and scientists. Elsevier; 2013.

282. Liu H, Shah S, Jiang W. On-line outlier detection and data cleaning. Computers & chemical engineering. 2004;28(9):1635–47.

283. Li Z, Zhang R, Lee CH, Lee YC. An evaluation of posture recognition based on intelligent rapid entire body assessment system for determining musculoskeletal disorders. Sensors. 2020;20(16):4414.

284. Yao W. Fine-Grained Hand Pose Estimation System based on Channel State Information. 2020;

285. Ghaleb FA, Kamat MB, Salleh M, Rohani MF, Abd Razak S. Two-stage motion artefact reduction algorithm for electrocardiogram using weighted adaptive noise cancelling and recursive Hampel filter. PloS one. 2018;13(11):e0207176.

286. Erer KS. Adaptive usage of the Butterworth digital filter. Journal of biomechanics. 2007;40(13):2934–43.

287. Pagnon D, Domalain M, Reveret L. Pose2Sim: An End-to-End Workflow for 3D Markerless Sports Kinematics—Part 1: Robustness. Sensors. 2021 Jan;21(19):6530.

288. Winter DA. Biomechanics and motor control of human movement. John Wiley & Sons; 2009.

289. Mehdizadeh S, Nabavi H, Sabo A, Arora T, Iaboni A, Taati B. Concurrent validity of human pose tracking in video for measuring gait parameters in older adults: a preliminary analysis with multiple trackers, viewing angles, and walking directions. Journal of NeuroEngineering and Rehabilitation. 2021;18(1):1–16.

290. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In 2017. p. 2223–32.

291. Bertalmio M, Sapiro G, Caselles V, Ballester C. Image inpainting. In 2000. p. 417–24.

292. Pratt AL, Ball C. What are we measuring? A critique of range of motion methods currently in use for Dupuytren's disease and recommendations for practice. BMC Musculoskelet Disord. 2016 Jan 13;17:20.

293. Giavarina D. Understanding bland altman analysis. Biochemia medica. 2015;25(2):141–51.

294. Reissner L, Fischer G, List R, Taylor WR, Giovanoli P, Calcagni M. Minimal detectable difference of the finger and wrist range of motion: comparison of goniometry and 3D motion analysis. J Orthop Surg Res. 2019 Dec;14(1):173.

295. Trejo RL, Ramirez MLG, Corral LEV, Marquez IR. Hand goniometric measurements using leap motion. In IEEE; 2017. p. 137–41.

296. Sandau M, Koblauch H, Moeslund TB, Aanæs H, Alkjær T, Simonsen EB. Markerless motion capture can provide reliable 3D gait kinematics in the sagittal and frontal plane. Medical engineering & physics. 2014;36(9):1168–75.

297. Li Y, Di H, Xin Y, Jiang X. Optical fiber data glove for hand posture capture. Optik. 2021 May 1;233:166603.

298. Pinto RF, Borges CD, Almeida A, Paula IC. Static hand gesture recognition based on convolutional neural networks. Journal of Electrical and Computer Engineering. 2019;2019.

299. Halim Z, Abbas G. A Kinect-Based Sign Language Hand Gesture Recognition System for Hearing- and Speech-Impaired: A Pilot Study of Pakistani Sign Language. Assistive Technology. 2015 Jan 2;27(1):34–43.

300. Metsis V, Jangyodsuk P, Athitsos V, Iversen M, Makedon F. Computer aided rehabilitation for patients with rheumatoid arthritis. In IEEE; 2013. p. 97–102.

301. Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale video classification with convolutional neural networks. In 2014. p. 1725–32.

302. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In Springer; 2018. p. 270–9.

303. Tammina S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. International Journal of Scientific and Research Publications (IJSRP). 2019;9(10):143–50.

304. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging. 2016;35(5):1285–98.

305. Mari A, Bromley TR, Izaac J, Schuld M, Killoran N. Transfer learning in hybrid classical-quantum neural networks. Quantum. 2020 Oct 9;4:340.

306. Jiang F, Zhang S, Wu S, Gao Y, Zhao D. Multi-layered gesture recognition with Kinect. J Mach Learn Res. 2015;16(1):227–54.

307. Panwar M, Singh Mehra P. Hand gesture recognition for human computer interaction. In: 2011 International Conference on Image Information Processing. 2011. p. 1–7.

308. Peng X, Wang L, Cai Z, Qiao Y. Action and gesture temporal spotting with super vector representation. In Springer; 2014. p. 518–27.

309. Chai X, Liu Z, Yin F, Liu Z, Chen X. Two streams recurrent neural networks for large-scale continuous gesture recognition. In IEEE; 2016. p. 31–6.

310. Camgoz NC, Hadfield S, Koller O, Bowden R. Using Convolutional 3D Neural Networks for User-independent continuous gesture recognition. In: 2016 23rd International Conference on Pattern Recognition (ICPR). 2016. p. 49–54.

311. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3d convolutional networks. In 2015. p. 4489–97.

312. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. In 2016. p. 265–83.

313. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC medical imaging. 2015;15(1):1–28.

314. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. Advances in neural information processing systems. 2017;30.

315. D'Eusanio A, Simoni A, Pini S, Borghi G, Vezzani R, Cucchiara R. A transformer-based network for dynamic hand gesture recognition. In IEEE; 2020. p. 623–32.

316. Joshi R, Gupta A. Performance Comparison of Simple Transformer and Res-CNN-BiLSTM for Cyberbullying Classification. arXiv preprint arXiv:220602206. 2022;

317. Schipper LG, Vermeer M, Kuper HH, Hoekstra MO, Haagsma CJ, Broeder AAD, et al. A tight control treatment strategy aiming for remission in early rheumatoid arthritis is more effective than usual care treatment in daily clinical practice: a study of two cohorts in the Dutch Rheumatoid Arthritis Monitoring registry. Annals of the Rheumatic Diseases. 2012 Jun 1;71(6):845–50.

318. Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R, et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. The Lancet. 2004 Jul 17;364(9430):263–9.

319. Schipper LG, van Hulst LTC, Grol R, van Riel PLCM, Hulscher MEJL, Fransen J. Meta-analysis of tight control strategies in rheumatoid arthritis: protocolized treatment has additional value with respect to the clinical outcome. Rheumatology (Oxford). 2010 Nov;49(11):2154–64.

320. Kay J, Upchurch KS. ACR/EULAR 2010 rheumatoid arthritis classification criteria. Rheumatology. 2012;51(suppl_6):vi5–9.

321. Davergne T, Rakotozafiarison A, Servy H, Gossec L. Wearable Activity Trackers in the Management of Rheumatic Diseases: Where Are We in 2020? Sensors (Basel). 2020 Aug 25;20(17):4797.

322. Song Y, Bernard L, Jorgensen C, Dusfour G, Pers YM. The Challenges of Telemedicine in Rheumatology. Frontiers in Medicine. 2021;1843.

323. Greenhalgh T, Vijayaraghavan S, Wherton J, Shaw S, Byrne E, Campbell-Richards D, et al. Virtual online consultations: advantages and limitations (VOCAL) study. BMJ open. 2016;6(1):e009388.

324. Weimers P, Ankersen DV, Munkholm P. Biomarkers for Remote Monitoring. In: Biomarkers in Inflammatory Bowel Diseases. Springer; 2019. p. 271–8.

325. Pope JE, Choy EH. C-reactive protein and implications in rheumatoid arthritis and associated comorbidities. In Elsevier; 2021. p. 219–29.

326. El Miedany Y, El Gaafary M, Youssef S, Bahlas S, Almedany S, Ahmed I, et al. Toward electronic health recording: evaluation of electronic patient-reported outcome measures system for remote monitoring of early rheumatoid arthritis. The Journal of rheumatology. 2016;43(12):2106–12.

327. Lester LE. The biomechanical analysis of the hand in rheumatoid arthritis patients with MCP arthroplasty [Internet] [m_rs]. University of Birmingham; 2009 [cited 2021 Jul 16]. Available from: https://etheses.bham.ac.uk/id/eprint/441/

328. Goldenberg DL, Bradley LA, Arnold LM, Glass JM, Clauw DJ. Understanding fibromyalgia and its related disorders. Prim Care Companion J Clin Psychiatry. 2008;10(2):133–44.

329. Wessel J. The effectiveness of hand exercises for persons with rheumatoid arthritis: a systematic review. Journal of Hand Therapy. 2004;17(2):174–80.

330. Yeager L. Effectiveness of adaptive utensils on hand and finger range of motion in individuals with rheumatoid arthritis. :22.

331. Depont F, Berenbaum F, Filippi J, Le Maitre M, Nataf H, Paul C, et al. Interventions to Improve Adherence in Patients with Immune-Mediated Inflammatory Disorders: A Systematic Review. PLoS One. 2015;10(12):e0145076.

332. Björk M, Thyberg I, Haglund L, Skogh T. Hand function in women and men with early rheumatoid arthritis. A prospective study over three years (the Swedish TIRA project). Scandinavian journal of rheumatology. 2006;35(1):15–9.

333. Mugii N, Hasegawa M, Matsushita T, Kondo M, Orito H, Yanaba K, et al. The efficacy of self-administered stretching for finger joint motion in Japanese patients with systemic sclerosis. The Journal of rheumatology. 2006;33(8):1586–92.

334. Rosique F, Losilla F, Navarro PJ. Applying Vision-Based Pose Estimation in a Telerehabilitation Application. Applied Sciences. 2021;11(19):9132.

335. Chua J, Ong LY, Leow MC. Telehealth using PoseNet-based system for in-home rehabilitation. Future Internet. 2021;13(7):173.

336. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC public health. 2020;20(1):1–9.

337. Cohen J. A coefficient of agreement for nominal scales. Educational and psychological measurement. 1960;20(1):37–46.

338. Singh JA, Saag KG, Bridges Jr SL, Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology guideline for the treatment of rheumatoid arthritis. Arthritis & rheumatology. 2016;68(1):1–26.

339. Mohammed RHA, Bhutta BS. Hand and Wrist Rheumatoid Arthritis. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 [cited 2022 Mar 31]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK560890/

340. Azhand A, Rabe S, Müller S, Sattler I, Heimann-Steinert A. Algorithm based on one monocular video delivers highly valid and reliable gait parameters. Sci Rep. 2021 Jul 7;11(1):14065.

341. Soto A, Camerino O, Iglesias X, Anguera MT, Castañer M. LINCE PLUS: Research software for behavior video analysis. Apunts Educació física i esports. 2019;3(137):149–53.

342. Symmons DP, Silman AJ. The Norfolk Arthritis Register (NOAR). Clin Exp Rheumatol. 2003 Oct;21(5 Suppl 31):S94-99.

343. Dale J, Paterson C, Tierney A, Ralston SH, Reid DM, Basu N, et al. The Scottish Early Rheumatoid Arthritis (SERA) Study: an inception cohort and biobank. BMC Musculoskeletal Disorders. 2016 Nov 9;17(1):461.

344. Norgeot B, Glicksberg BS, Trupin L, Lituiev D, Gianfrancesco M, Oskotsky B, et al. Assessment of a Deep Learning Model Based on Electronic Health Record Data to Forecast Clinical Outcomes in Patients With Rheumatoid Arthritis. JAMA Netw Open. 2019 Mar 1;2(3):e190606.

345. An G, Akiba M, Omodaka K, Nakazawa T, Yokota H. Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images. Scientific reports. 2021 Mar 1;11(1):1-9.

346. Creswell A, White T, Dumoulin V, Arulkumaran K, Sengupta B, Bharath AA. Generative adversarial networks: An overview. IEEE Signal Processing Magazine. 2018;35(1):53–65.

Appendices

A.1 Influence of markers on Openpose: CycleGAN and image inpainting approaches

OpenPose accuracy was tested in Chapter 4 to evaluate the effect of visible markers on participant's skin when using the keypoint detection methodology. To this end, two different image-to-image translation techniques were adopted. These techniques were applied to get a fully annotated dataset and avoid the visibility of the markers on the hand data recorded using the optical motion capture system. This step aims to capture the features of one frame and understand, with training, how these features can be converted to another image.

First was the CycleGAN (290) process. The CycleGAN process is part of a class of networks known as Generative Adversarial Networks (GAN) (346), often used for translating imagery data content. The used GANs are a class of deep learning that generate images that look realistic, when instead are fake. Given images with and without markers in input to the network, one part generates candidates (generator) and the other evaluates them (discriminator). This approach aims to overcome the lack of large-scale annotated datasets and the way in which data are pre-processed to provide ground truth for kinematic estimation. In this work, the CycleGAN process (Figure 52) was tested to generate an image-to-image mapping function, as shown between visible and not visible markers on the hand.

The CycleGAN approach can perform automatic unsupervised training utilizing the frames from one data source and converting It to a different source. This image-to-image translation process made use of both images with and without markers to suppress the visible markers from the caption images. Therefore, two small subsets of video recordings were collected during the study described in Chapter 4, one with marker and the other without, both containing 62 frames. The outcomes from the marker removal utilizing the CycleGAN methodology resulted in blurred images Figure 53 (second row), and another methodology was therefore investigated.

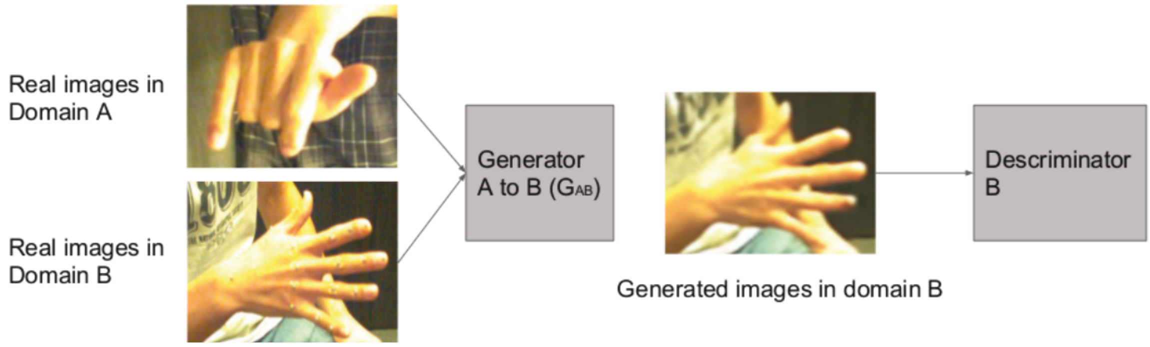


Figure 52: The CycleGAN process used to test the effect of markers on the output from OpenPose.

Image inpainting (291) was also utilized with the purpose of removing visible marker location. This approach, unlike the CycleGAN, made use only of the frames with visible markers, without the need for the images without markers. The results in output were less blurred compared with the CycleGAN system, as illustrated in Figure 53 (third row).

Figure 53: Results of image inpainting and CycleGAN Techniques. Top row are original images where the reflective markers can be seen, the central row are results from CycleGAN, the bottom row are results from image inpainting; note that with the inpainting process the markers are not visible and the images are less blurred than in the CycleGAN Process.

One sequence contained the frames with markers, the other sequence contained the frames in output from the image inpainting technique (without the visible markers). the results keypoint detection model was run on the two video sequences producing the same results in output. The overall chosen final adopted process is shown in the flowchart in Figure 54.

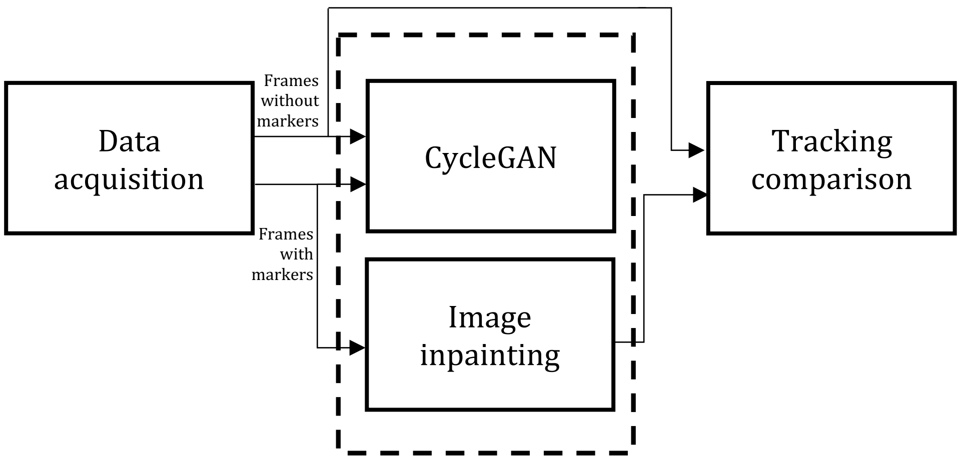


Figure 54: Flowchart of the final development of both the acquisition and the analysis used in this investigation using image inpainting; starting from the motion capture trials of 10 healthy volunteers to the development of the final hand tracking model. The inpainting process make uses of just images with markers to suppress the visible markers from the caption images. This approach produced better quality images than the CycleGAN for marker removal.

Finally, to check that the output from OpenPose stayed the same of the same frames with and without makers, proving therefore that no variation was present due to the probability variation within OpenPose, the model was executed and compared. The results show the tracking stayed the same in the frames with and without makers. The executed code showing matching outputs is pasted below.

```
for i,x in enumerate(coordinates):
  print(f"{i}) Image coordinates output matching: {x.sort()==coordinates[0].sort()}")

    0) Image coordinates output matching: True
    1) Image coordinates output matching: True
    2) Image coordinates output matching: True
    3) Image coordinates output matching: True
    4) Image coordinates output matching: True
    5) Image coordinates output matching: True
    6) Image coordinates output matching: True
    7) Image coordinates output matching: True
    8) Image coordinates output matching: True
```

A.2 Letters of support

National Rheumatoid Arthritis Society

Dear Msk MEC Accelerator Steering Committee,

As a Deputy CEO of the National Rheumatoid Arthritis Society (NRAS) I am pleased to write this letter in support of the 'AI-powered Rheumatic Assessment (AIRA)' project. NRAS is the only patient led organisation in the UK with a specific focus on rheumatoid arthritis (RA) and juvenile idiopathic arthritis (JIA). As such we are fascinated by the project Arthronica and will be supporting Letizia Gionfrida in some initial focus groups with people living with RA.

My understanding is that the AIRA project will further improve the readiness of the device for use with rheumatology patients and put it in a position for use in a larger clinical trial, that NRAS would also be willing to support.

I strongly recommend this project for: its potential to be transformational especially for those being diagnosed and who require regular monitoring to get their disease under control quickly so as to prevent long term damage, pain and links to other comorbidities which in turn lead to negative impact on all aspects of a person's life. I understand that this product is answering to a clinical need for targeted treatment strategies in a cost-effective way that can increase productivity and reduce the need for high-cost second line monitoring solutions.

We trust you will consider the funding application favourably and the cost of NRAS' involvement in patient recruitment and input to be factored in at £400. We are delighted to commit to patient involvement at every stage of this research.

Sincerely,

Clare Jacklin
Deputy CEO

Ground Floor, 4 The Switchback, Gardner Road, Maidenhead, Berkshire, SL6 7RJ
Private company limited by guarantee. Registered in England & Wales Number 7127101 Registered Charity Numbers 1134859 SC039721
www.nras.org.uk
Freephone Helpline 0800 298 7650 General Tel: 0845 458 3969/01628 823524 Fax: 0845 458 3971 Email enquiries@nras.org.uk

Faculty of Life Sciences
and Medicine

School of Immunology and Microbial Sciences

Department of Inflammation Biology
Centre for Rheumatic Diseases

3.48, WEC
Denmark Hill Campus
10 Cutcombe Road
London SE5 9RJ
Telephone 020 3 299 7772
Email james.galloway@kcl.ac.uk

Friday, 05 July 2019

Dear MSK MEC Accelerator Steering Committee,

As a Senior Clinical Lecturer at King's College London and honorary consultant in Rheumatology at King's College Hospital, I am pleased to write this letter in support of the 'AI-powered Rheumatic Assessment (AIRA)' project. I am interested in the hand scanner for rheumatoid arthritis that is being developed by Letizia Gionfrida and Arthronica and would be supportive of evaluating it in my clinic. The AIRA project will improve the readiness of the device for use with rheumatology patients and put it in a position for use in a larger clinical trial, that I would willingly support.

I strongly recommend this project for: its high translational potential and the societal impact that it could have on people with rheumatoid arthritis, particularly people with early inflammatory rheumatoid arthritis, who require frequent interventions to prevent chronic pain and disability. This product is answering to a clinical need for targeted treatment strategies in a cost-effective way that can increase productivity and reduce the need for high-cost second line monitoring solutions.

Yours sincerely,

Dr James Galloway
Senior lecturer, King's College London
Honorary consultant in rheumatology, King's College Hospital

A.3 Table of permissions granted for figures

Figures	Copyright	Permission granted
Figure 1	Copyright © 2017 Nanayakkara, Cotugno, Vitzilaios, Venetsanos, Nanayakkara and Sahinkaya. This is an open-access article distributed under the terms of the Creative Commons Attribution license.	Reproduced with permission from (58), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 3	Copyright © 2013 IEEE.	Reprinted, with permission, from (89), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 5 (A)	Copyright © 2013 IEEE.	Reproduced with permission from (124), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 5 (B)	Copyright © 2013 Weichert, Bachmann, Rudak, Fisseler.	Reproduced with permission from (123), licensee MDPI, Basel, Switzerland, distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 6	Copyright © 2019 Machine Learning Mastery.	Reproduced with permission granted via the Machine Learning Mastery contact form https://machinelearningmastery.com/contact/
Figure 7	Copyright © 2015 Elsevier.	Reproduced with permission from (144), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 8	Copyright © 2013 IEEE.	Reproduced with permission from (145), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 9	Copyright © 2017 IEEE.	Reproduced with permission from (147), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 10	Copyright © 2017 IEEE.	Reproduced with permission from (16), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 11	Copyright © 2017 IEEE.	Reproduced with permission from (152), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 13	Copyright © 2021 Wang.	Reproduced with permission from (25), licensee MDPI, Basel, Switzerland, distributed

		under the terms of the Creative Commons (CC BY) Attribution license.
Figure 14	Copyright © 2021 Wang.	Reproduced with permission from (25), licensee MDPI, Basel, Switzerland, distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 15	Copyright © 2007 National Institutes of Health.	Reproduced with permission from the National Institutes of Health. distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 16	Copyright © 2012 Clinical and Experimental Rheumatology.	Reproduced with permission from (215), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 17	Copyright © 1994 PubMed Central (PMC).	Reproduced with permission from (216).
Figure 18	Copyright © 2016 by The Journal of Rheumatology Publishing Co. Ltd.	Reproduced with permission from (229), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 19	Copyright © 2015 Williams et al. This is an Open Access publication distributed under the terms of the Creative Commons Attribution licence.	Reproduced with permission from (251), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 20	Copyright © 1996 British Journal of Rheumatology	Reproduced with permission from (269), distributed by Oxford University press on behalf of the British Society for Rheumatology.
Figure 24	Copyright © 2013 IEEE.	Reproduced with permission from (280), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 26	Copyright © 2017 Hamberger & Weiss LLP.	Reproduced with permission by Hamberger & Weiss LLP granted via email in October 2021.
Figure 27	Copyright © 2017 Hamberger & Weiss LLP.	Reproduced with permission by Hamberger & Weiss LLP granted via email in October 2021.
Figure 28	Copyright © 2017 Hamberger & Weiss LLP.	Reproduced with permission by Hamberger & Weiss LLP in October 2021.
Figure 35	Copyright © 2021 Quantum.	Reproduced with permission from (305), distributed under the terms of the Creative Commons (CC BY) Attribution license.

Figure 40	Copyright © 2019 IEEE.	Reproduced with permission from (157), distributed under the terms of the Creative Commons (CC BY) Attribution license.
Figure 47	Copyright © 2022 by DNN Corp and/or its suppliers. All rights reserved.	Reproduced with permission from The British Society for Rheumatology, distributed under the terms of the Creative Commons (CC BY) Attribution license.